Residential Land Values in the Washington, DC Metro Area: New Insights from Big Data*

Morris A. Davis
Rutgers University

Stephen D. Oliner
American Enterprise Institute and UCLA

Edward J. Pinto
American Enterprise Institute

Sankar Bokka
FNC, Inc.

January 21, 2016

Abstract

We use new property-level data to estimate the price of land from 2000 to 2013 for nearly the universe of detached single-family homes in the Washington, DC metro area and characterize the housing boom-bust cycle in land and house prices at a fine geography. The data show that land prices were more volatile than house prices everywhere, but especially so in the areas where land was inexpensive in 2000. We demonstrate that the change in the land share of house value during the boom was a significant predictor of the decline in house prices during the bust, highlighting the value of focusing on land in assessing house-price risk.

JEL Classification Numbers: C55, E32, R14, R23, R31, R32
Keywords: Land, Housing, House Prices, Housing Boom and Bust, Financial Crisis

*The views expressed are ours alone and do not represent those of the institutions with which we are affiliated. We received expert research assistance from Rahee Jung, Jessica Li, Lilla Lukacs, Eleanor Qian, Shuyi Yu and especially Urbashee Paul. In addition, we received helpful comments from Norrine Brydon, Robert Dorsey, Gabriel Ehrlich, Stuart Gabriel, Richard Green, Stuart Rosenthal, Joe Tracy, Stijn Van Nieuwerburgh, Mike Viehweg, Susan Wachter, and participants at the International Conference on Collateral Risk (March 2014), the AREUEA National Conference (May 2014) and the Third and Fourth Annual International Conferences on Housing Risk (September 2014 and October 2015). We thank FNC, Inc. and Marshall & Swift, a CoreLogic company, for providing the data for the paper. We also thank the UCLA Rosalinde and Arthur Gilbert Program in Real Estate, Finance and Urban Economics for generous funding.
Corresponding author: Morris A. Davis, morris.a.davis@rutgers.edu

1 Introduction

One of the key lessons from the housing boom and bust, the foreclosure surge, and the financial crisis is that participants in housing and mortgage markets, along with policymakers, need better information to evaluate the risks in these markets. With the right data and tools in place, we believe that future volatility in housing and mortgage markets can be tempered through sound lending practices and appropriate regulation. The research here contributes to this objective by quantifying collateral risk at a fine level of geography. Many studies (Davis and Heathcote (2007), Davis and Palumbo (2008), Haughwout, Orr, and Bedoll (2008), Kok, Monkkonen, and Quigley (2014) and Nichols, Oliner, and Mulhall (2013), among others) have shown that most of the risk in the price of a house reflects risk to the value of the underlying land. In this paper, we estimate the price of residential land from 2000 to 2013 for 740,000 detached single-family homes in the Washington, DC metropolitan area. These properties represent nearly the universe of all such homes in the area under study. To our knowledge, no previous study has estimated land values for individual parcels on such a large scale.

As emphasized by Davis and Heathcote (2007), Davis and Palumbo (2008), and Nichols, Oliner, and Mulhall (2013), housing can be viewed as a bundle consisting of a structure that provides shelter and land that provides utility because of its particular location. Housing structures are relatively easy to reproduce using a known technology whereas the supply of land ready for development is inelastically supplied. If the supply curve for structures is flat, and the supply curve for land is steep, then shocks to the demand for housing should result primarily in an increase in the quantity of structures built and an increase in the price of land.

Although housing can be affected by both supply shocks and demand shocks, the available evidence suggests that demand shocks account for most of the time-series variation in house prices and housing market activity. Well-calibrated models with only supply shocks (Davis and Heathcote (2005), Favilukis, Ludvigson, and Van Nieuwerburgh (2015) and Garriga, Manuelli, and Peralta-Alva (2012), for example) have trouble replicating the observed volatility of house prices. In addition, a casual look at aggregate data strongly supports the notion that demand shocks drive housing markets. Figure 1 graphs the four-quarter percent change in real house prices against single-family housing starts, both for the U.S. as a whole. The two series are very highly correlated. Figure 1 is consistent with the idea that housing markets are best characterized as subject to large and persistent demand shocks that simultaneously boost prices and quantities. Given its low elasticity of supply, land will account for the bulk of the price variation in the housing bundle.

We estimate land prices with a method that takes advantage of the rich information in a new dataset of home prices and construction costs that we have assembled. Briefly summarizing this method, we start by measuring the implied value of land for newly-built homes as the difference between the observed sale price and the estimated construction cost for the new structure. Since we focus on new homes, this method does not require an estimate of the depreciation of existing housing structures – a significant benefit given the paucity of such information. The resulting land prices for newly-built homes provide a market-based "stake in the ground" for a specific time period, which we apply (with some adjustments we discuss below) to all existing homes in the same zip code. Finally, we create property-level time series for home value, land value, and structure value that incorporate the stake in the ground and that reflect the changes over time in indices of house prices and construction costs for the property's local area.

Our results for the cross-sectional pattern of land prices and land use conform in broad terms to the predictions of a standard urban model, i.e. Muth (1969) and Mills (1972). In the Washington, DC metro area, the price of land tends to be less expensive and lot sizes tend to increase the farther parcels are located from the city center. At the same time, we find that the Washington, DC area cannot be characterized by a uniform downward land-price gradient from the city center, as the price of land is relatively low in close-in areas populated by less affluent residents.

When we focus on time-series variation in land prices, our results are quite striking. In the places where land was cheap in 2000, largely the outer suburbs and closer-in areas with limited locational amenities, the price of land jumped more than 600 percent during the boom from 2000 to 2006. In contrast, in the affluent inner suburbs where land was expensive in 2000, its price increased less than 200 percent over the same period. We see the same pattern in reverse during the bust. The price of land declined only about 10 percent from its peak in expensive areas but plunged nearly 75 percent in initially cheap areas. Although land prices were everywhere more volatile than house prices, the price of land was most volatile in areas where land was inexpensive and represented a small fraction of home value in 2000.

Unlike land prices, house prices in the Washington area rose by roughly the same percentage in most places from 2000 to 2006. Importantly, this implies a simple analysis of house prices would have provided few clues to mortgage underwriters or other housing-market participants as to which areas were most vulnerable to a possible downturn. After the peak, house prices and land prices moved down in tandem, with steep price drops where land was initially cheap and much milder declines in places where land was expensive. Our analysis provides a key insight for future assessments of house-price risk. An explosive rise in land prices, even when not accompanied by a relatively large increase in house prices, may signal an increased risk of a severe house-price decline.

Our final result relates to the land-leverage hypothesis of Bostic, Longhofer, and Redfearn (2007). This hypothesis predicts that house prices will be most volatile in areas where the value of land accounts for much of the value of housing, i.e. areas with the highest land share of property value. The prediction will hold if land prices are everywhere more volatile than house prices and if the ups and downs in land prices are similar in the various areas under study. Our results are consistent with the first condition but not the second, as we find that the land-price cycle was much sharper in the areas with low land shares. Consequently, in the Washington, DC area, the drop in house prices after 2006 was the most severe in areas with low land shares - the opposite of the prediction of the land leverage hypothesis. We believe our paper, which permits the measurement of land prices and land shares with fine geographic detail, is the first study to provide evidence that is inconsistent with the hypothesis.

The rest of the paper is organized as follows. The next section discusses the unique, property-level dataset for the Washington, DC area that we create for the analysis. Section 3 describes in detail how we estimate land prices for the individual properties in the dataset. Section 4 presents a snapshot of the cross-sectional variation in the Washington metropolitan area, while section 5 lays out the time-series results for the recent boom-bust cycle. Section 6 concludes.

2 Data

2.1 Geographic Coverage

Our dataset covers the city of Washington, DC and the other primary jurisdictions in the Washington, DC metropolitan area: Montgomery and Prince George's Counties in Maryland, along with Arlington, Fairfax, Prince William and Loudoun Counties and the city of Alexandria in Virginia. Table 1 presents basic information about the eight jurisdictions in the dataset. Fairfax County and Montgomery County are the most populous areas, with each having slightly more than a million residents as of 2014. The eight jurisdictions have a total population of about five million people. Compared with the United States as a whole, the eight jurisdictions on average are quite affluent, have a well-educated population, and are ethnically diverse.

However, there are some significant differences across the jurisdictions. Arlington, Fairfax and Loudoun counties have a combination of high median income, high educational attainment, and a relatively small share of the population that is Black or Hispanic. In contrast, Prince George's County has considerably lower median income, a much smaller of adults with a college education, and a population that is almost 80 percent Black or Hispanic. The City of Alexandria, Montgomery and Prince William Counties and Washington, DC have a blend of the characteristics at these extremes.

2.2 Source Data

The core of our dataset consists of property-level information for detached single-family homes in the eight jurisdictions. The primary data source is the National Collateral Database (NCD) produced by the mortgage technology company, FNC, Inc. As described in Dorsey, Hu, Mayer, and Wang (2010), the NCD covers virtually all residential properties throughout the United States, blending data from public records and home appraisals. The property-level information includes physical characteristics, location, the history of sale prices, appraised value, and the latest tax assessment. Importantly, the NCD file shows the year in which the home was built, its sale price when new, and its lot size. This information enables us to estimate land value with the methodology described in section 3.

We merge the NCD file with two other pieces of information for each home. The first is FNC's proprietary estimate of the home's market value as of either 2013:Q3 or 2014:Q3.[1] FNC maintains a set of automated valuation models (AVMs) and selects the model that performs best for a given location.[2] For example, in Montgomery County, the preferred AVM keys off the home's tax assessment, while in Prince William County, the preferred model is a hedonic regression estimate.

The other piece of information merged with the NCD record is the estimate of the home's reconstruction cost as of 2013:Q3 or 2014:Q2 from Marshall & Swift, a CoreLogic company.[3] Marshall & Swift's construction cost estimates are used extensively by property insurance companies, building contractors, appraisers, and government agencies. Their estimate of reconstruction cost represents the full cost of rebuilding the home from the ground up as a new structure. This includes the cost of materials and labor, equipment rentals, builder and subcontractor profit margins, permits and fees, and all applicable taxes associated with

[1]The earlier date applies to properties in Fairfax, Montgomery, Prince George's, and Prince William Counties, as well as Washington, DC, while the latter date applies to properties in the other jurisdictions. The dates differ because we obtained the FNC data in two tranches. The differing valuation dates do not pose a problem because we use zip-level house price indices, as described below, to adjust estimated market values across the sample period.

[2]For further information, see "FNC Automated Valuation Models," April 2014 (*http://www.fncinc.com/*). This document is available from FNC on request.

[3]As with the AVMs, we received the Marshall & Swift reconstruction cost estimates in two tranches.

residential construction in the local area. The estimated reconstruction cost differs across homes based on a number of factors, but the main drivers are location, total square footage of living space, number of stories, and whether the home is attached or detached from other housing units.

As noted above, each home's estimated market value and reconstruction cost pertain to a single quarter around the end of our sample period. We create property-level time series for both variables with local house price indices and construction cost indices. The house-price indices are those produced by FNC using a spatial hedonic model estimated with the NCD data (see Dorsey, Hu, Mayer, and Wang (2010) for details). FNC constructs the indices monthly back to January 2000 at various levels of geography, including the five-digit zip-code level. For each home, we use the index for the five-digit zip code in which it is located.[4] The construction cost indices come from Marshall & Swift. These indices measure the cost of constructing several different models of homes in a given area, incorporating the various costs described above. Marshall & Swift produces the indices monthly for three-digit zip codes throughout the country. The level of aggregation is higher than for the house price indices because construction costs do not vary within narrowly defined areas. We use the construction cost index for the three-digit zip code in which each home is located.

2.3 Property-Level Dataset

To create the dataset for our empirical work, we merged the three sources of property-level data and then removed properties in the NCD file that could not be address-matched to the Marshall & Swift file or that did not have an AVM. Among the remaining properties, we applied a series of data-quality screens.

The first of these screens removed properties in the NCD file that likely are either condominiums or townhouses. In an initial pass through the data, we eliminated properties that Marshall & Swift classified as condos or end-unit townhouses.[5] Then, to screen out other attached units, we dropped the properties with very small lots.[6]

Three additional screens removed properties with evident data errors or properties that

[4]The indexes for individual metropolitan areas and higher-level aggregations are posted at *http://www.fncrpi.com/*. The zip-level indexes are available from FNC on request.

[5]The Marshall & Swift file only identifies townhouses that are end units. It does not distinguish other townhouses from detached homes.

[6]Specifically, we removed properties with lots in the bottom 1/2 percent of the jurisdiction's lot size distribution. We then manually checked a random sample of properties that remained after this trim to look for the presence of townhouses. Based on this manual check, we set a minimum lot size for each jurisdiction so that the properties with lot sizes slightly above the threshold were predominantly detached single-family homes. The resulting minimum lot size ranged from 3,500 square feet in Washington, DC to 6,000 square feet in Loudoun County, with a median across the eight jurisdictions of 4,000 square feet.

were outliers. Specifically, we dropped properties that were listed as having the same address as another property, properties with extremely large lots (in the top 1/2 percent for their jurisdiction), and properties for which the difference between the AVM and the estimated construction cost was in the bottom 1 percent or top 1 percent for their zip code. We did the latter trim at the zip level to allow for systematic differences in land values.[7]

Table 2 shows the size of the resulting dataset, together with the 2013 Census estimate of the number of detached single-family homes in each jurisdiction. Our dataset contains a total of 740,000 properties, implying a coverage rate of 97.5 percent relative to the Census estimate. Hence, our dataset contains nearly the universe of detached single-family homes in the Washington, DC area. For the eight jurisdictions individually, the coverage rates range from about 82 percent in Alexandria and Washington, DC to more than 100 percent in Fairfax and Prince William Counties.[8]

2.4 Zip Codes and Zip Groups

The number of properties in each five-digit zip code varies widely. Of the 192 five-digit zips in the dataset, several have more than 10,000 properties, while others have just a handful. These differences reflect the underlying variation in the total number of residential properties by zip code and the split of these properties between detached single-family homes and other types of housing.

To deal with this variation, we attached the zip codes with sparse data to an adjacent zip code (or in some cases, more than one adjacent zip) to create geographic units with at least 1000 properties. This aggregation condensed the original 192 zip codes into 141 geographic units that we refer to as "zip groups." About three-quarters of the zip groups consist of a single zip code, and most of the others combine two zips. Appendix table A.1 lists all the individual zip codes in the dataset and the created zip groups, showing the number of properties in each zip group, the jurisdiction in which it is located and information related to the estimation of land value.

[7] For zip codes with relatively few properties, we trimmed at the "zip group" level. We discuss the distinction between zip codes and zip groups below.

[8] Coverage rates above 100 percent can arise for two reasons. First, the Census Bureau estimates come from the 2013 American Community Survey and are subject to sampling error. The Census Bureau estimates that the 90 percent confidence band for Fairfax County is ± 1.8 percent around the reported count, while that for Prince William County is ± 3.1 percent; the 95 percent and 99 percent bands would be considerably wider. Second, our dataset includes some townhouses, which causes our totals to slightly overstate the number of detached single-family homes.

3 Estimating Land Value

We estimate land value for individual homes in the Washington, DC area with an approach has its roots in traditional land valuation practices – see Babcock (1932) for an early reference – and that blends the methods used in two strands of the modern literature. One strand has relied on observed land sales; recent examples include Albouy and Ehrlich (2013a), Albouy and Ehrlich (2013b), Kok, Monkkonen, and Quigley (2014) and Nichols, Oliner, and Mulhall (2013). This approach has the virtue of incorporating observed market prices. However, the volume of land transactions in the older, established parts of a metropolitan area is generally too sparse to estimate changes in land values over time for specific localities. Reflecting this limitation, Nichols, Oliner, and Mulhall (2013) estimated an aggregate land price index for each of 23 metropolitan areas, but did not attempt to calculate indexes for different parts of each metro area. An alternative approach (see, for example, (Davis and Palumbo, 2008)) measures land value indirectly as the difference between the value of a representative house and an estimate of the depreciated reconstruction cost of the structure on the lot. This method can be implemented with less data than the transaction-based approach. But calculating land value as a residual means that the estimates inherit the measurement errors elsewhere in the accounting framework.

Our method takes advantage of the strengths of each approach, while circumventing their weaknesses. Specifically, we use market prices to estimate land value for individual homes at a given time and then compute internally consistent measures of home value and structure value. This provides a market-based stake in the ground for a specific time period. We then create property-level time series for home value, land value, and structure value that incorporate the stake in the ground and that reflect the changes over time in indices of house prices and construction costs for the property's zip group. Importantly, this method does not require any estimate of depreciation for housing structures - a significant benefit given the absence of information on depreciation for local housing markets.

3.1 Land, Home and Structure Value in a Single Reference Period

3.1.1 Land Value for Newly Built Homes

We take advantage of two key pieces of property-level information: The sales price for a house when it was first built and the estimated cost of rebuilding the structure as new.[9] Let $h^n_{i,z,r}$ denote the sale price of new home i located in zip group z and built in period r,

[9] Throughout the paper, we use the naming convention that a "house" is a combination of "land" and "structure."

and let $s^n_{i,z,T}$ denote the reconstruction cost for the structure if built new in period T. We distinguish between the date of the estimate of the replacement cost of the structure, T, and the date the new home was built and sold, r, because the estimated reconstruction cost provided to us from Marshall & Swift is as of 2013:Q3 or 2014:Q2 for all properties, whereas the sale prices are almost entirely from periods before 2013.

We first move the estimated reconstruction cost of specific home i in zip group z back to the period of the initial home sale using the Marshall & Swift index of construction costs. Denote this index value for any arbitrary period t as $p^c_{z,t}$. We compute the reconstruction cost back to period r, the date of the initial home sale, as

$$s^n_{i,z,r} \quad = \quad s^n_{i,z,T} \cdot \left(\frac{p^c_{z,r}}{p^c_{z,T}} \right) \tag{1}$$

In words, the reconstruction cost in period r (the year the structure was built) is assumed to equal the reconstruction cost in period T times the ratio of the construction cost index in r to the index value in T. Given this estimate, and the (standard) assumption that the value of housing is equal to the sum of the reconstruction cost of the structure and the market value of the land, we can compute the value of land when the home was first built, $l^n_{i,z,r}$, as the residual

$$l^n_{i,z,r} \quad = \quad h^n_{i,z,r} - s^n_{i,z,r} \tag{2}$$

For this method to work, the estimated reconstruction cost in period T must reflect the same structure as when the house was built. Homes that have had a major renovation or expansion fail this requirement and thus cannot be used in equation (2). Because our dataset does not provide the history of improvements for a property, we assume that homes built in 2000 or after have not had major structural changes by 2013. Thus, the earliest period for which we use equation 2 is 2000:Q1.

Two other considerations define the sales used in equation (2). First, we employ only arms-length sales of non-distressed property to help ensure that the implied land values will be applicable to the full population of properties. And second, we include sales that occur not only in the year the home was built, but also in the year after, as some new homes may sit on the market for a while before being sold.

An implicit assumption behind equation (2) is that reconstruction cost serves as a good proxy for the unobserved market value of the new housing structure. We believe this is a reasonable assumption. The equality of reconstruction cost and structure value is a basic equilibrium condition in a housing market with ongoing construction activity by profitmaximizing builders (see Rosenthal (1999) and Schulz and Werwatz (2011), for example). Although this condition will not hold exactly at all times, builders have an incentive to adjust the level of construction activity to close gaps that open up. The limited available evidence (Rosenthal, 1999) suggests that the adjustment occurs fairly rapidly. Accordingly, using reconstruction cost as a proxy for the market value of a new structure should not generate significant measurement error.

The transactions in our dataset include some sales of vacant land in addition to sales of finished homes, but the two types of sales are not identified explicitly. It is important to distinguish new home sales from land sales because equation (2) is only applicable to the former; for land transactions, the sale price itself measures the property's land value. We classify the sales in our dataset as follows. Let $\widehat{l}^n_{i,z,r}$ denote the estimate of land value from blindly applying equation (2) without knowledge of whether the sale involves a finished home or land. If $\widehat{l}^n_{i,z,r}$ is positive and greater than 10 percent of the sale price $h^n_{i,z,r}$, we assume this is a new home sale and use equation (2) to value the land. Conversely, if $\widehat{l}^n_{i,z,r}$ is negative and greater than 10 percent of the sale price in absolute value, we assume the reported sale is a land sale (since the sale price is significantly below the cost of the structure that ultimately appears on the site); in such cases, we set the land value equal to the observed sale price. Finally, when $\widehat{l}^n_{i,z,r}$ is within 10 percent of the sale price in absolute value, we do not classify the transaction as either a land or new home sale and set it aside.[10] After applying this procedure, about 11 percent of the properties in the full dataset have an estimated land value.

The number of parcels for which we have an estimate of $l^n_{i,z,r}$ differs widely by zip group. Not surprisingly, there are many observations in outlying areas where the housing stock is relatively new and far fewer observations in closer-in areas that were almost completely developed before 2000. As shown in the "# Sales-based land values" column of appendix table A.1, six of the ten zip groups in Washington, DC have fewer than 25 estimates of land value over the entire 2000-2013 period, while most zip groups in Prince William County have more than 1,000. Although more observations clearly would be preferred to fewer, even a small number of land values - if reasonably representative - can provide the required stake

[10] We also exclude duplicate sales for a given property and sales with apparent data errors. Specifically, we drop transactions classified as land sales that occur in the year after the house was built (which is logically impossible), transactions for which the sale price is more than five times greater than the property's AVM or less than 10 percent of the AVM, and properties with multiple transactions when the highest imputed land value is more than five times greater than the lowest imputed value (true changes of that magnitude are highly unlikely given that they would have had to occur within a window that included, at most, the year in which the house was built and the following year). In addition, we exclude properties with multiple sales records on a single date when the lowest and highest sale prices differ by more than $5,000; when the prices differed by less than $5,000, we use the first sales record.

in the ground for our methodology.

3.1.2 Land Value for All Homes

The next step is to use the market-based land values for new homes in zip group z to impute land values for all homes in that zip group. One approach would be to regress the market-based land values on lot size, time, and possibly other variables, using the regression coefficients to compute fitted land values for other houses in the zip group. The difficulty, however, is that some zip groups have too few observations to produce reliable coefficients from such a regression.

In light of this issue, we use an alternative approach. To begin, we designate a "reference quarter," denoted using the time subscript R, for each zip group. The reference quarter is the period in which we place the stake in the ground. If the initial year in our analysis, 2000, contains at least ten market-based land values for a given zip group, we define the reference quarter to be the median quarter (in 2000) of the included sales. Most zip groups have more than the minimum of ten sales in 2000, often many times more. For other zip groups, we add years one at a time until we have at least ten market-based land values or until we reach 2006, whichever comes first. The reference quarter in this case is the median quarter of the included land values across the included years.[11]

For each zip group, we then calculate the median land value per square foot and median lot size for the same sale properties used to define the reference quarter.[12] These two variables are the key inputs for estimating land value for all properties in the reference quarter. We use medians to reduce the influence of outliers, and we separate total land value into the value per square foot and lot size in order to account for the established finding that land value per square foot declines with lot size.[13]

Our procedure is as follows. Let $\bar{\ell}^n_{z,R}$ denote the median land price per square foot of the newly-built sale properties used to determine the reference quarter R for zip group z,

[11]We set 2006 as the final year to ensure that every zip group has a reference quarter as close to the beginning of the sample period as possible. For all but a handful of zip groups, the final year used to determine the reference quarter is 2004 or earlier, so the 2006 cutoff seldom comes into play.

[12]In 38 of the zip groups, we calculate the reference quarter, the median land value per square foot, and the median lot size based solely on the presumed home sales, omitting the presumed land sales. We do this because the land values from these latter sales tend to be substantially higher than the land values from the presumed home sales. In all likelihood, these presumed land sales are actually home sales that we misclassified when using the allocation rule described above. The column in appendix table A.1 labeled "Type of sales used" indicates the zip groups for which we use only the presumed home sales.

[13]This is the so-called "plattage effect" that has been documented in many previous empirical studies of land prices. In our dataset, larger residential lots sell for less per square foot than otherwise identical smaller lots subject to the same zoning because the extra square footage cannot be used to build an additional house. Discussions of the lack of proportionality between square footage and lot value date back as far as Babcock (1932), Bernard (1913), Hurd (1903) and Mertzke (1927).

and let $\bar{q}^n_{z,R}$ denote the median lot size of those properties. In addition, for the full set of properties in zip group z, let $q_{j,z}$ denote the lot size of the jth property in zip group z (no time subscript is needed because lot size for any individual property is constant). We estimate an "adjustment function" f, described below, to take account of the effect on land price per square foot that arises because the median lot size for the newly-built sale properties differs from the lot size for any individual property j in the zip group. With this notation, the estimated land value for property j in zip group z in reference period R is

$$l_{j,z,R} \equiv \left[\bar{\ell}^n_{z,R} \cdot f\left(q_{j,z}, \bar{q}^n_{z,R}\right)\right] q_{j,z} \tag{3}$$

where the term in brackets is the estimated land value per square foot after adjusting for the lot-size effect.

To calculate f, we pool the sale properties across all the zip groups within a given jurisdiction and regress the natural log of the land price per square foot for each property on a set of zip group dummy variables, a third-order polynomial in time, and the natural log of lot size. The dummy variables control for differences in the average level of land prices across zip groups, while the polynomial in time controls for the cycle in land prices. We estimate the regression separately for each jurisdiction to allow for differences in the size of the plattage effect, as reflected in the estimated coefficient on the natural log of lot size. This coefficient, denoted by α, determines the adjustment factor for each property:

$$f\left(q_{j,z}, \bar{q}^n_{z,R}\right) = \exp\left[\alpha \ln\left(q_{j,z}\right) - \alpha \ln\left(\bar{q}^n_{z,R}\right)\right] = \left(\frac{q_{j,z}}{\bar{q}^n_{z,R}}\right)^{\alpha} \tag{4}$$

where the middle expression is exponentiated to convert the fitted land price per square foot from natural logs to levels.

The upper panel of figure 2 plots the estimated adjustment factor for Montgomery County for lots ranging in size from 5,000 square feet to 45,000 square feet (roughly one acre), relative to a baseline quarter-acre lot (10,890 square feet).[14] As shown, large lots have a much lower price per square foot than small lots, consistent with the findings in the literature. For example, the price per square foot for a half-acre lot (21,780 square feet) is less than 60 percent of that for the baseline quarter-acre lot. The lower panel shows the implied variation in total lot value (lot size times land price per square foot) as lot size varies. The estimated premium for additional square footage in Montgomery County is modest - only about 10 percent for a half-acre lot relative to an otherwise identical quarter-acre lot. This small

[14]The adjustment functions for the other jurisdictions are similar to that for Montgomery County, as the estimates of α across the eight jurisdictions are bunched in a tight range from -0.80 to -0.88.

premium reflects the fact that the extra land cannot be used to build another house.

3.1.3 Home and Structure Value for All Homes

Given the property-level estimate of land value in the reference quarter from equation (3), we complete the picture for each property by estimating its total market value (land plus structure) and the structure value alone in the reference quarter. Recall that the dataset includes an AVM estimate for any home j's market value in a specific quarter in either 2013 or 2014. Denote this estimate as $h_{j,z,T}$. We move that estimate back to the reference quarter with the FNC house price index for the home's zip group. That is, denoting the FNC house price index for the home's zip group z for any period t as $p^h_{z,t}$, we compute the expected market value of any home j in reference period R as

$$h_{j,z,R} \quad = \quad h_{j,z,T} \cdot \left(\frac{p^h_{z,R}}{p^h_{z,T}} \right) \tag{5}$$

Given this estimate of the market value of home j in reference period R, and given the estimate of the value of the land for home j in the reference period, we can estimate the depreciated structure value for each home in the reference period as

$$s_{j,z,R} \quad = \quad h_{j,z,R} - l_{j,z,R} \tag{6}$$

3.2 Time Series of Land, Home, and Structure Values

The property-level estimates of total home value, land value, and structure value in the reference period R provide the stake in the ground from which we can compute property-level time series from 2000:Q1 to 2013:Q4.

We create an estimate of total home value for home j in zip z in each period t from 2000:Q1 through 2013:Q4 as the home value in the reference period times the ratio of the home price index in period t to the home price index in reference period R, i.e.

$$h_{j,z,t} \quad = \quad h_{j,z,R} \cdot \left(\frac{p^h_{z,t}}{p^h_{z,R}} \right) \tag{7}$$

Similarly, we create an estimate of depreciated structure value of home j in zip group z for each period t from 2000:Q1 through 2013:Q4 as

$$s_{j,z,t} \quad = \quad s_{j,z,R} \cdot \left(\frac{p^c_{z,t}}{p^c_{z,R}} \right) \tag{8}$$

where, as mentioned, p^c is the Marshall & Swift construction cost index for zip group z. We then compute the value of land in each period as the residual:

$$l_{j,z,t} = h_{j,z,t} - s_{j,z,t} \tag{9}$$

These property-level time series are the building blocks for the analysis that follows. We aggregate across the properties in a zip group to obtain time series for average home, land, and structure values, along with the average land share of home value. Obviously, our property-level estimates are subject to measurement errors: for example, not every house in the same zip group will appreciate at the same rate. All that we require for the analysis that follows, however, is that the zip-level aggregates are unbiased, which we expect to occur as long as our property-level errors reflect classical measurement error.

As a check, we examine the implied land shares for the zip groups. For the large majority of zip groups, the land shares present no obvious issues. However, in some zip groups, the estimated land share in 2000:Q1 is close to zero or even negative, before rising over time, while in other zip groups, the land share exceeded 90 percent in some periods. Anomalous values such as these could arise if the sale properties - from which the market-based land values are determined - are not representative of the full set of properties in the zip group. This could occur, for example, if the homes built since 2000 are located in parts of the zip group for which land prices are generally higher (or lower) than average.

For these groups, we adjust the starting value for the land shares in 2000:Q1. Specifically, any shares below five percent are adjusted to be exactly five percent, while those greater than 60 percent are lowered to that value.[15] These adjustments nudge the 2000:Q1 land shares toward reasonable values. For each property in an affected zip group, we multiply the adjusted 2000:Q1 land share by the original estimate of home value in 2000:Q1 to produce an adjusted estimates of land value and structure value (as home value less land value).[16] With this revised stake in the ground for 2000:Q1, we re-create the time series for each property using equations (7), (8) and (9).

[15] Note that Davis and Palumbo (2008) make similar adjustments in their data set. There were also three zip groups with no estimated land share because of an absence of new home sales. For these zip groups, we estimate the 2000:Q1 land share as the fitted value from a regression – estimated with data for all the other zip groups – of the 2000:Q1 land share on a constant and the 2000:Q1 average AVM value.

[16] Appendix table A.1 identifies the 41 zip groups (of 141 in total) for which these adjustments were made. The time-series results presented in section 5 are essentially the same whether we include or exclude the 41 zip groups from the analysis. This can be seen by comparing figures 10-14 to appendix figures A.1-A.5.

4 Cross-Sectional Patterns

This section presents a snapshot of the variation across the zip groups for house prices, land values, and other property characteristics. We use a series of heat maps that divide the zip groups into quintiles, using data for 2013:Q4, the latest period in our dataset. In each map, the darkest shade represents the quintile with the highest values for each variable and the lightest shade represents the lowest quintile.

Figure 3 shows the pattern of house prices in 2013:Q4 across the zip groups, where the house price for each zip group is the average AVM for the homes in that group. As shown, house prices are the highest in Alexandria, Arlington County, Fairfax County, the southwestern part of Montgomery County (Bethesda, Chevy Chase, and Potomac), and the part of Washington, DC that borders Arlington and Montgomery Counties. These areas include the most sought-after addresses in the Washington area. House prices are lower in outlying areas and in locales with relatively low income and generally high proportions of Black and Hispanic residents (Prince George's County and the adjoining parts of Montgomery County and Washington, DC). The spread of average AVM values across the zip groups is wide, ranging from less than $315,000 for every zip group in the lowest quintile to as much as $1.7 million in the highest quintile.

Figure 4 presents the parallel distribution of average lot values, calculated as described in section 3. As can be seen by comparing figures 3 and 4, the geographic pattern for lot values is quite close to that for home prices. Moreover, the range of average lot values across zip groups is extremely wide, from as little as $24,000 to more than $1.2 million. Clearly, house prices are high in some places and lower in others largely because of differences in the value of the underlying land.

As indicated in figure 5, lots tend to be relatively small in close-in areas and larger further out. None of the zip groups in Washington, DC proper has an average lot size greater than 9,000 square feet, while zip groups in the exurbs have lots that average as much as nearly four acres. This pattern is consistent with the canonical urban model of Muth (1969) and Mills (1972) which predicts that households respond to the higher price of land closer to the city center by economizing on land use.[17]

Given the wide differences in average lot size across the zip groups, measured land prices per square foot will be influenced by the plattage effect described in section 3. That is, large lots will have a lower value per square foot, all else equal, simply because much of the square footage is above and beyond that needed to build a house at that location. To control for

[17]Technically, the prediction of the model relates to the size of the housing structure relative to the lot, rather than to lot size per se (see McMillen (2006)). The heat map for the land-to-structure ratio (not shown) displays qualitatively the same pattern as figure 5.

this effect, we calculated the average land price per square foot for a quarter-acre lot in each zip group, using the adjustment function in the top panel of figure 2 for Montgomery County and parallel adjustment functions for other jurisdictions. This standardized land price does a better job than the raw price per square foot of capturing the amenity value of land in various locations.

As shown in figure 6, the spatial pattern for the quarter-acre land price is very similar to that for house prices from figure 3. This high correlation confirms that differences in location-specific amenity value are an important driver of the variation in house prices across zip groups.

Finally, figure 7 portrays the average land share of property value across the zip groups in 2013:Q4. Comparing figures 6 and 7 shows that the land share tends to be high in the zip groups with high land prices and vice versa. Notably, nearly all the zip groups in Arlington County, Alexandria, the affluent part of Washington, DC and the close-in parts of Fairfax and Montgomery Counties have land shares in the highest quintile or the next quintile. In contrast, most of the zip groups in Prince George's county are in the lowest or second-lowest quintile. This difference in land shares will play a central role in the analysis of the housing cycle in the next section.[18]

5 The Price Cycle Since 2000

5.1 House Prices

We use the zip-level hedonic price indexes published by FNC, Inc. to examine the appreciation in house prices from 2000 and 2006 and the subsequent decline from 2006 to 2012. The heat maps in figures 8 and 9 portray the respective periods.

Both figures show substantial variation across the zip groups, consistent with the heterogeneity found in other recent studies of house prices using data at a fine level of geography as in Ferreira and Gyourko (2011), Guerrieri, Hartley, and Hurst (2013) and Mian and Sufi (2009). From 2000 to 2006, prices in the lowest quintile of zip groups rose 89 to 127 percent, compared with a range that tops 200 percent in the highest quintile. Then, from 2006 to 2012, prices in the quintile with the largest declines fell 40 to 51 percent, while prices

[18]Despite this general pattern, some zip groups have land shares that differ considerably from those in neighboring zip groups. More often than not, these zip groups are ones for which we adjusted the land shares from the raw values produced by the estimation procedure. These zips groups include 20720 and 20721 in Prince George's County, the combination of 22134 and 22172 in Prince William County, and 20120 in Fairfax and Loudoun Counties, all of which have land shares (even after adjustment) above those in adjacent zips. The land shares for these zip groups likely are less reliable than for those for other zip groups.

dropped less than 12 percent in the quintile with the smallest declines. Ten zip groups in this latter quintile actually saw prices rise on net over 2006-12. Moreover, one of those zips (20815, in Montgomery County) is located less than five miles from another zip in the same county (20902) where prices dropped 34 percent. Clearly, the Washington, DC area cannot be characterized as a single, homogeneous housing market.

Interestingly, the house price declines shown in figure 9 have a pronounced geographic pattern. Prices fell the least in affluent, close-in areas while the steepest declines were concentrated in the outlying areas and in Prince George's County. That is, the exurbs and places with large Black and Hispanic populations were hit the hardest during the bust.

These two seemingly distinct areas share an important characteristic: they both have relatively low land prices (as we saw in figure 6). Thus, the places in the metropolitan area with cheap land experienced the most severe house-price crash.[19] This pattern suggests that land could play a key role in house-price dynamics, a connection we explore in the rest of this section.

5.2 Land Prices, House Prices and Construction Costs

We begin by examining the movement in land prices over our full sample period. To focus on the connection between the level of land prices and the magnitude of the price cycle, we aggregate the 141 zip groups into quintiles based on the quarter-acre land price in 2000:Q1.

Figure 10 presents the resulting time series for land prices. The figure demonstrates that the amplitude of the price cycle was systematically related to the initial level of land prices. The price swing - both the rise and then the decline - was the greatest for the zip groups where land was initially the cheapest. As one moves from the lowest quintile of land prices to each higher quintile, the price cycle becomes progressively milder.[20]

To provide perspective on the size of the swing in land prices, figures 11 and 12 compare the percent changes in land prices, house prices, and construction costs for the zip groups aggregated into the same land-price quintiles as in figure 10. Figure 11 presents the compar-

[19] Figure 6 displays land prices in 2013:Q4, the latest period in our dataset. The use of 2013:Q4 could raise concerns that the magnitude of the post-2006 house price decline mechanically influenced the zip-group distribution of land prices shown in figure 6, making the correlation uninteresting. However, as we will show, the same correlation emerges between the level of land prices in 2000:Q1 (the initial period in the dataset) and the post-2006 house price decline. Hence, the connection between the level of land prices and the severity of the house price decline is not an artifact of using land prices as of 2013:Q4.

[20] This result is the land-price counterpart to the pattern shown for house prices in the Case-Shiller indexes for separate price tiers, available at *http://us.spindices.com/index-family/real-estate/sp-case-shiller*. For each of the 20 included metropolitan areas, the magnitude of the post-2000 house price increase and the subsequent decline was greater for low-price homes than for high-price homes. Guerrieri, Hartley, and Hurst (2013) document the same pattern over 2000-06 for a broader set of metropolitan areas.

ison for the boom phase of the cycle (2000-06), while figure 12 covers the bust (2006-12). In both figures, house prices are measured by the FNC zip-level indices, while the construction cost indices are those produced by Marshall & Swift.

Three important results are immediately evident from figure 11. First, during the boom phase of the cycle, land prices rose much more than house prices in every quintile, while construction costs increased only modestly. Hence, the rise in house prices over 2000-06 largely reflected the appreciation in land value. Second, house prices increased by similar amounts in each of the quintiles. This implies that monitoring the rise in house prices in each of the land-price quintiles would have provided little information about the magnitude of the post-2006 price drop, which we know from figure 10 was especially large in the quintiles with low land prices. Third, unlike the nearly uniform rise in house prices, the increase in land prices was exceptionally large in the lowest land-price quintile, at more than 600 percent, and then moderated from quintile to quintile (though it remained large even in the highest quintile). Consequently, land prices contain information about potential overvaluation beyond that in house prices themselves.

Turning to figure 12, what we observe after the 2006 peak mirrors in some respects what happened during the run-up through 2006. Land prices fell more than house prices in every quintile, with staggering losses in the lowest price quintile, where land prices plunged more than 70 percent. Because construction costs continued to rise over 2006-12, the drop in land prices more than accounted for the decline in house prices, emphasizing the central role of land in house-price swings. One difference, however, from figure 11 is that the magnitude of the house-price drop varies across the quintiles, and markedly so. Whereas house prices retreated only 10 percent on net in the zip groups in the highest land-price quintile, they fell more than 35 percent in the lowest quintile. Differences of this magnitude likely would be associated with sharp differences in the performance of the underlying mortgage loans, highlighting the potential value of information from land price movements in assessing risk.

5.3 Land Share of Property Value

Another key difference across the quintiles relates to the land share of property value. Figure 13 plots the time series of the average land share in each land-price quintile. Figure 13 shows that the land share is low in the locations where land is cheap and high where land is expensive. For example, in 2000, the land share ranged from about 20 percent in the lowest land-price quintile to more than 50 percent in the highest two quintiles. At the end of the sample period in 2013, the gap was even a bit wider.

The difference in land shares across the quintiles explains arithmetically why land prices rose so much more than house prices over 2000-06 in the lowest quintiles. With land representing a small share of house value, the accounting relationship that connects house prices to land and structure prices could only hold if land prices had risen much more than house prices. In the quintiles with higher land shares, the same accounting relationship would hold with a smaller rise in land prices relative to house prices.

Our results concerning land shares have important implications for the land-leverage hypothesis of Bostic, Longhofer, and Redfearn (2007). This hypothesis states that house prices will be more volatile, all else equal, in places where land represents a relatively large share of property value. Bostic, Longhofer, and Redfearn (2007) found that the hypothesis was supported in their detailed study of house prices in Wichita, Kansas. A necessary, but not sufficient, condition for the hypothesis to be valid is that land prices have wider swings than structures prices (so that land prices are more volatile than house prices). The findings from numerous previous studies have satisfied this condition.[21] Our results for the Washington, DC area do as well, as we found that land prices rose more and then fell more than house prices in every land-price quintile.

Nonetheless, our results do not support the central prediction of the hypothesis. That is, we do not find that house-price volatility is the greatest in the zip groups with the highest land shares. Rather, we see the opposite pattern, as shown in table 3. The zip-group quintile with the lowest land prices and the lowest land shares experienced the most severe drop in house prices after 2006. Conversely, the two quintiles with the highest land shares had the mildest post-2006 decline in house prices. During the 2000-06 boom, there were only small differences in house price appreciation across the quintiles, but the slight differences go against the land-leverage hypothesis, with house prices rising the least in the quintile with the highest land share in 2000.

Our results do not support the land-leverage hypothesis because land prices were so much more volatile in the zip groups with low land shares than in the zip groups with high shares. Evidently, land prices in the areas with low land shares had a larger speculative component than land prices elsewhere.

5.4 Predictive Power of Changes in Land Share

Reflecting these differential movements in land prices, the land share has been more volatile since 2000 in the zip groups with initially cheap land than in the zip groups with

[21] See Davis and Heathcote (2007), Davis and Palumbo (2008), Haughwout, Orr, and Bedoll (2008), Nichols, Oliner, and Mulhall (2013) and Sirmans and Slade (2012) for supporting evidence in U.S. metropolitan areas. For international evidence, see Bourassa, Haurin, Haurin, Hoesli, and Sun (2009), Bourassa, Hoesli, Scognamiglio, and Zhang (2011), Schulz and Werwatz (2011) and Wu, Gyourko, and Deng (2012).

more expensive land (recall figure 13). An important question is whether the amount by which the land share rose during the boom helps predict the magnitude of the house price decline during the bust. If so, real-time monitoring of land prices and land shares would be useful for market participants and policymakers.

To address this question, figure 14 plots the percent change in house prices from 2006 to 2012 against the rise in the land share in percentage points from 2000 to its maximum value in any year through 2006.[22] Each diamond represents one of the 141 zip groups. The figure shows that the rise in the land share during the boom did, in fact, have predictive power for the subsequent drop in house prices. The regression line fit through the scatter plot has a highly significant slope coefficient of -0.86, implying that an additional 10 percentage point rise in the land share was associated with a 8.6 percentage point drop in house prices after the peak.[23]

To interpret this result, note that a rise in the land share indicates that the gap between house prices and construction costs has widened. Large increases in this gap proved to be unsustainable over the recent housing cycle. Eventually, house prices reverted to fundamentals, bringing down the land share of property value.

6 Conclusions

We have provided a detailed picture of the recent boom-bust cycle in house prices and land prices for the Washington, DC metropolitan area, using an unprecedented dataset that covers close to the universe of detached single-family homes in the area under study. The rich property-level information allows us to estimate land value for the 740,000 homes in the dataset. To our knowledge, no previous research has estimated residential land values at the property level on such a large scale.

The paper yields important new facts about the recent housing cycle in the Washington area. First, the swing in house prices and land prices varied widely across locations. The cycle was mildest in the affluent, close-in parts of the metro area. It was considerably greater in the more distant suburbs and in areas with a large Black or Hispanic population – places where land is relatively cheap. Second, land prices were more volatile than house prices

[22] We use this definition of the rise in the land share because for some zip groups the land share peaked in 2005 rather than 2006.

[23] Note that the rise in the land share retains its predictive power even after controlling for the rise in house prices from 2000 to 2006. In a regression of the change in house prices over 2006-12 on the change in house prices over 2000-06 and the maximum rise in the land share over 2000-06, the coefficient on the land share is significant at the 1 percent level. Moreover, the R^2 of this regression is 0.33, compared with only 0.16 when the change in the land share is excluded.

everywhere, but especially so in the areas with initially inexpensive land. This held true even for different locations in relatively close proximity. Third, changes in the land share of property value were a useful predictor of subsequent change in house prices: the areas with the largest increases in the land share tended to suffer the sharpest drop in house prices during the bust. The predictive power of this single variable is especially notable because it works across a large, diverse metropolitan area subject to a variety of shocks. These results highlight the value of focusing on land for assessing house-price risk and suggest that market participants and policymakers should be particularly attentive to situations that involve a rapid appreciation of land prices from initially low levels.

Our results cut against the land-leverage hypothesis, which holds that house prices will be more volatile in areas where land represents a large share of property value. Previous research has supported the hypothesis when assessed across cities, but our study is the first to examine it within a large, diverse metropolitan area. We find the opposite of what the hypothesis predicts. House prices were most volatile in the parts of the metropolitan area with low land shares because the land price swing in those areas was much wider than elsewhere.

References

Albouy, D., and G. Ehrlich (2013a): "The Distribution of Urban Land Values: Evidence from Market Transactions," University of Illinois Working Paper. 7

——— (2013b): "Metropolitan Land Values and Housing Productivity," University of Illinois Working Paper. 7

Babcock, F. M. (1932): *The Valuation of Real Estate*. McGraw-Hill. 7, 10

Bernard, A. D. (1913): *Some Principles and Problems of Real Estate Valuation*. United States Fidelity and Guaranty Company. 10

Bostic, R. W., S. D. Longhofer, and C. L. Redfearn (2007): "Land Leverage: Decomposing Home Price Dynamics," *Real Estate Economics*, 35(2), 183–208. 3, 18

Bourassa, S. C., D. R. Haurin, J. L. Haurin, M. Hoesli, and J. Sun (2009): "House Price Changes and Idiosyncratic Risk: The Impact of Property Characteristics," *Real Estate Economics*, 37(2), 259–278. 18

Bourassa, S. C., M. Hoesli, D. Scognamiglio, and S. Zhang (2011): "Land Leverage and House Prices," *Regional Science and Urban Economics*, 41(2), 134–144. 18

Davis, M. A., and J. Heathcote (2005): "Housing and the Business Cycle," *International Economic Review*, 46(3), 751–784. 1

——— (2007): "The Price and Quantity of Residential Land in the United States," *Journal of Monetary Economics*, 54(8), 2595–2620. 1, 18

Davis, M. A., and M. G. Palumbo (2008): "The Price of Residential Land in Large U.S. Cities," *Journal of Urban Economics*, 63(1), 352–384. 1, 7, 13, 18

Dorsey, R. E., H. Hu, W. J. Mayer, and H.-c. Wang (2010): "Hedonic Versus Repeat-Sales Housing Price Indexes for Measuring the Recent Boom-Bust Cycle," *Journal of Housing Economics*, 19(2), 75–93. 4, 5

Favilukis, J., S. C. Ludvigson, and S. Van Nieuwerburgh (2015): "The Macroeconomic Effects of Housing Wealth, Housing Finance and Limited Risk-Sharing in General Equilibrium," *Journal of Political Economy*, Forthcoming. 1

Ferreira, F., and J. Gyourko (2011): "Anatomy of the Beginning of the Housing Boom: U.S. Neighborhoods and Metropolitan Areas, 1993-2009," NBER Working Paper 17374. 15

Garriga, C., R. Manuelli, and A. Peralta-Alva (2012): "A Model of Price Swings in the Housing Market," Federal Reserve Bank of St. Louis Working Paper 2012-022A. 1

Guerrieri, V., D. Hartley, and E. Hurst (2013): "Endogenous Gentrification and House Price Dynamics," *Journal of Public Economics*, 100(C), 45–60. 15, 16

Haughwout, A. F., J. A. Orr, and D. Bedoll (2008): "The Price of Land in the New York Metropolitan Area," Discussion Paper 14, Federal Reserve Bank of New York, Current Issues in Economics and Finance. 1, 18

Hurd, R. M. (1903): *Principles of City Land Values*. Real Estate Record Association. 10

Kok, N., P. Monkkonen, and J. M. Quigley (2014): "Land Use Regulations and the Value of Land and Housing: An Intra-Metropolitan Analysis," *Journal of Urban Economics*, 81, 136–148. 1, 7

McMillen, D. P. (2006): "Testing for Monocentricity," in *A Companion to Urban Economics*, ed. by R. J. Arnott, and D. P. McMillen. 14

Mertzke, A. J. (1927): "Real Estate Appraising," National Association of Real Estate Boards. 10

Mian, A., and A. Sufi (2009): "The Consequences of Mortgage Credit Expansion: Evidence from the U.S. Mortgage Default Crisis," *Quarterly Journal of Economics*, 124(4), 1449–96. 15

Mills, E. S. (1972): *Studies in the Structure of the Urban Economy*. Johns Hopkins Press. 2, 14

Muth, R. F. (1969): *Cities and Housing: The Spatial Pattern of Urban Residential Land Use*. University of Chicago Press. 2, 14

Nichols, J. B., S. D. Oliner, and M. R. Mulhall (2013): "Swings in Commercial and Residential Land Prices in the United States," *Journal of Urban Economics*, 73(1), 57–76. 1, 7, 18

Rosenthal, S. S. (1999): "Residential Buildings and the Cost of Construction: New Evidence on the Efficiency of the Housing Market," *The Review of Economics and Statistics*, 81(2), 288–302. 9

Schulz, R., and A. Werwatz (2011): "Is There an Equilibrating Relationship between House Prices and Replacement Cost? Empirical Evidence from Berlin," *Journal of Urban Economics*, 69(3), 288–302. 9, 18

Sirmans, C. F., and B. A. Slade (2012): "National Transaction-based Land Price Indices," *The Journal of Real Estate Finance and Economics*, 45(4), 829–845. 18

Wu, J., J. Gyourko, and Y. Deng (2012): "Evaluating Conditions in Major Chinese Housing Markets," *Regional Science and Urban Economics*, 42(3), 531–543. 18

Table 1: Population and Demographic Characteristics in 2014

County/city	Population	Median household income	Percent bachelors degree	Percent Black or Hispanic
Arlington Co., VA	226,908	$109,266	71.5	23.8
City of Alexandria, VA	150,575	$86,809	62.8	37.3
Fairfax Co., VA	1,137,538	$110,674	60.3	25.6
Loudoun Co., VA	363,050	$122,294	58.7	20.4
Montgomery Co., MD	1,030,447	$97,765	58.5	36.1
Prince George's Co., MD	904,430	$72,290	31.0	78.9
Prince William Co., VA	446,094	$92,104	39.2	41.3
Washington, DC	658,893	$71,648	55.0	58.1
Memo: U.S. total	318,857,056	$53,046	30.1	29.7

Note: The definition of Black or Hispanic is Hispanic or Latino of any race plus Black or African American alone and not Hispanic or Latino. Median household income is in 2013 dollars.

Sources: U.S. Census Bureau, 2014 American Community Survey, Tables DP02 (percent with bachelors degree), DP03 (median household income), and DP05 (population and percent Black or Hispanic).

Table 2: Count of Detached Single-family Homes

County/city	Dataset	Census	Percent coverage
Arlington Co., VA	27,184	28,436	95.6
City of Alexandria, VA	8,724	10,673	81.7
Fairfax Co., VA	194,186	193,704	100.2
Loudoun Co., VA	57,884	61,903	93.5
Montgomery Co., MD	175,918	183,395	95.9
Prince George's Co., MD	163,534	167,121	97.9
Prince William Co., VA	83,276	78,439	106.2
Washington, DC	29,591	35,886	82.5
Total	**740,297**	**759,557**	**97.5**

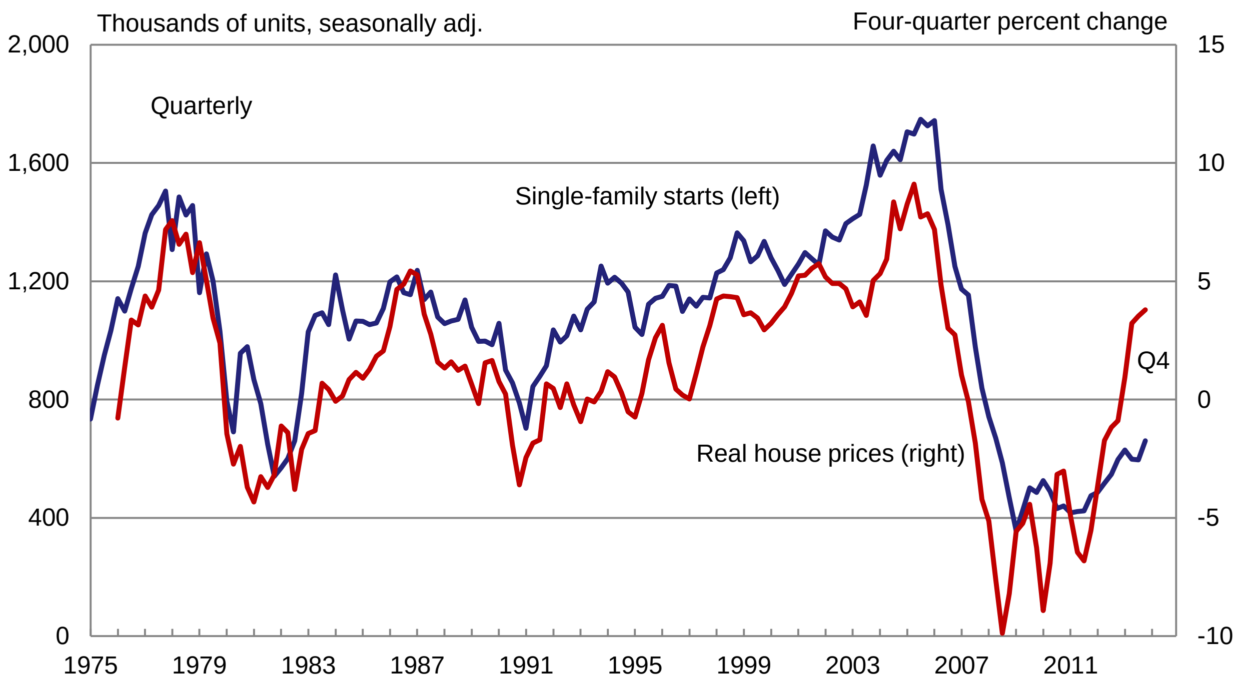
Sources: Authors' calculations based on dataset created from data provided by FNC, Inc. and Marshall & Swift, a CoreLogic company; U.S. Census Bureau, 2013 American Community Survey, one-year estimates, Table DP04.

Table 3: Land Shares and Changes in House Prices by Zip-Group Quintiles

Quintiles, quarter-acre land price, 2000:Q1	Land share (pct., annual average)			House-price change (pct.)	
	2000	2006	2012	2000-06	2006-12
Lowest	19.4	57.2	25.6	142.4	-36.0
Second	32.0	64.9	42.2	150.2	-33.3
Third	37.8	66.2	47.0	139.8	-28.9
Fourth	51.0	74.4	63.6	149.3	-21.0
Highest	54.5	73.1	66.5	121.6	-9.6

Sources: Authors' calculations based on dataset created from data provided by FNC, Inc. and Marshall & Swift, a CoreLogic company.

Figure 1: Single-family Housing Starts and Real House Prices, 1975:1 - 2013:4

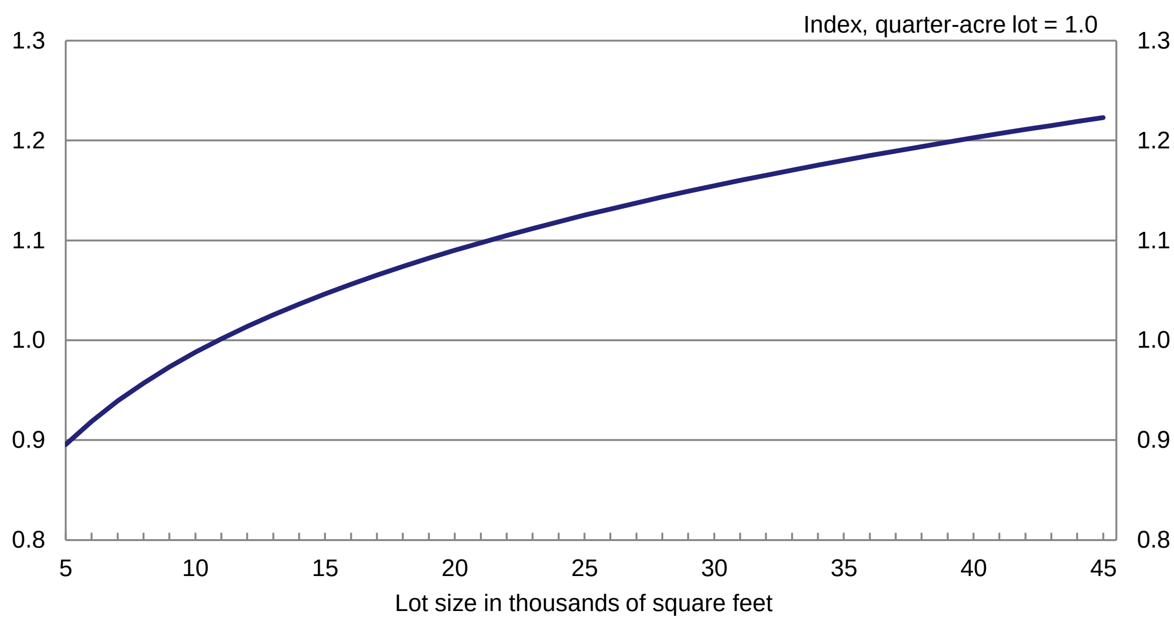


Note: Real house prices are measured as the FHFA all-transactions house price index divided by the price index for personal consumption expenditures.

Sources: Single-family housing starts, Census Bureau; real house prices, FHFA and Bureau of Economic Analysis.

Figure 2: Estimated Effect of Lot Size on Land Price per Square Foot and Lot Value in Montgomery County, MD

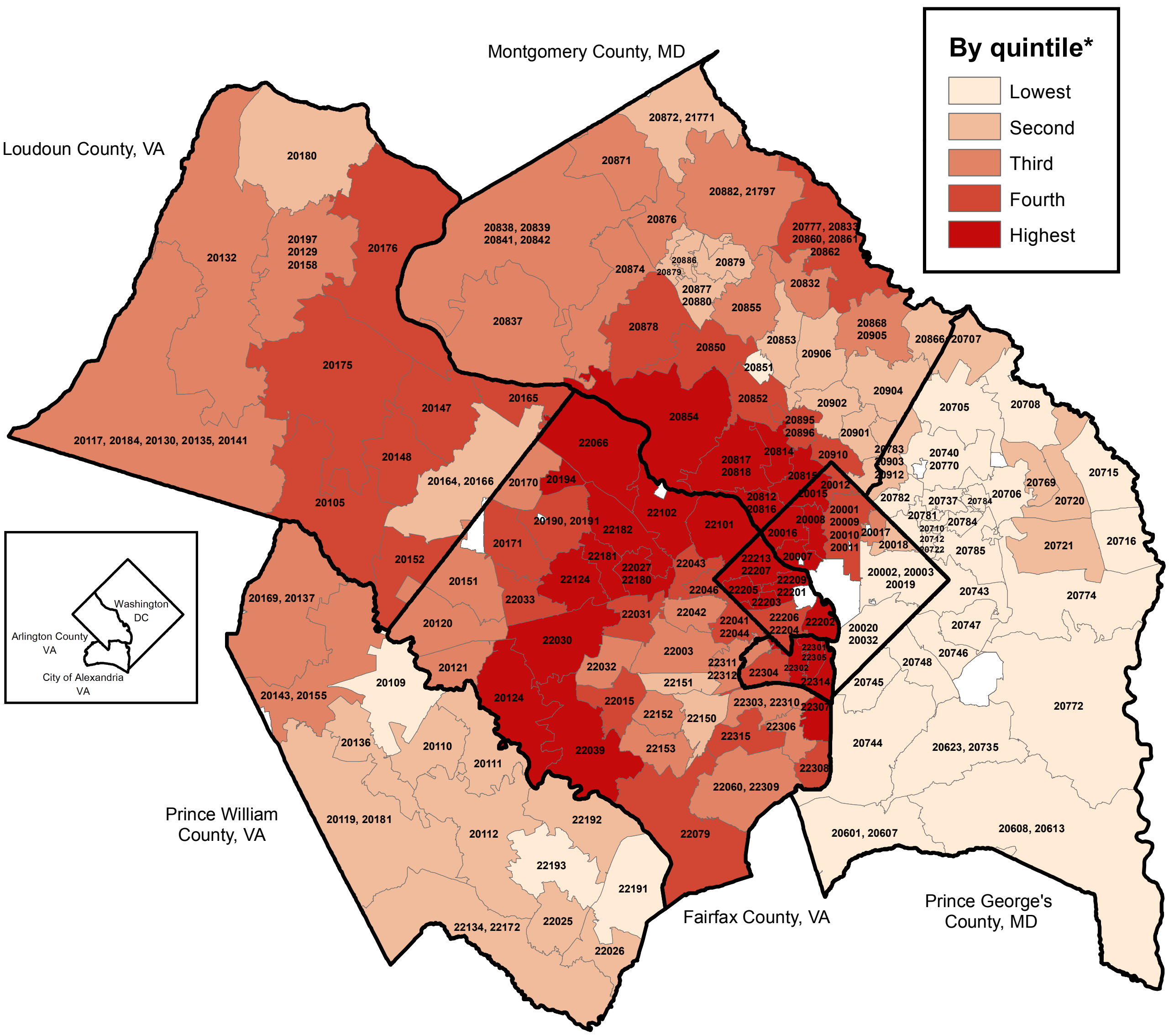
(a) Land Price per Square Foot



(b) Lot Value

Source: Authors' calculations using data from FNC, Inc. and Marshall & Swift, a CoreLogic company.

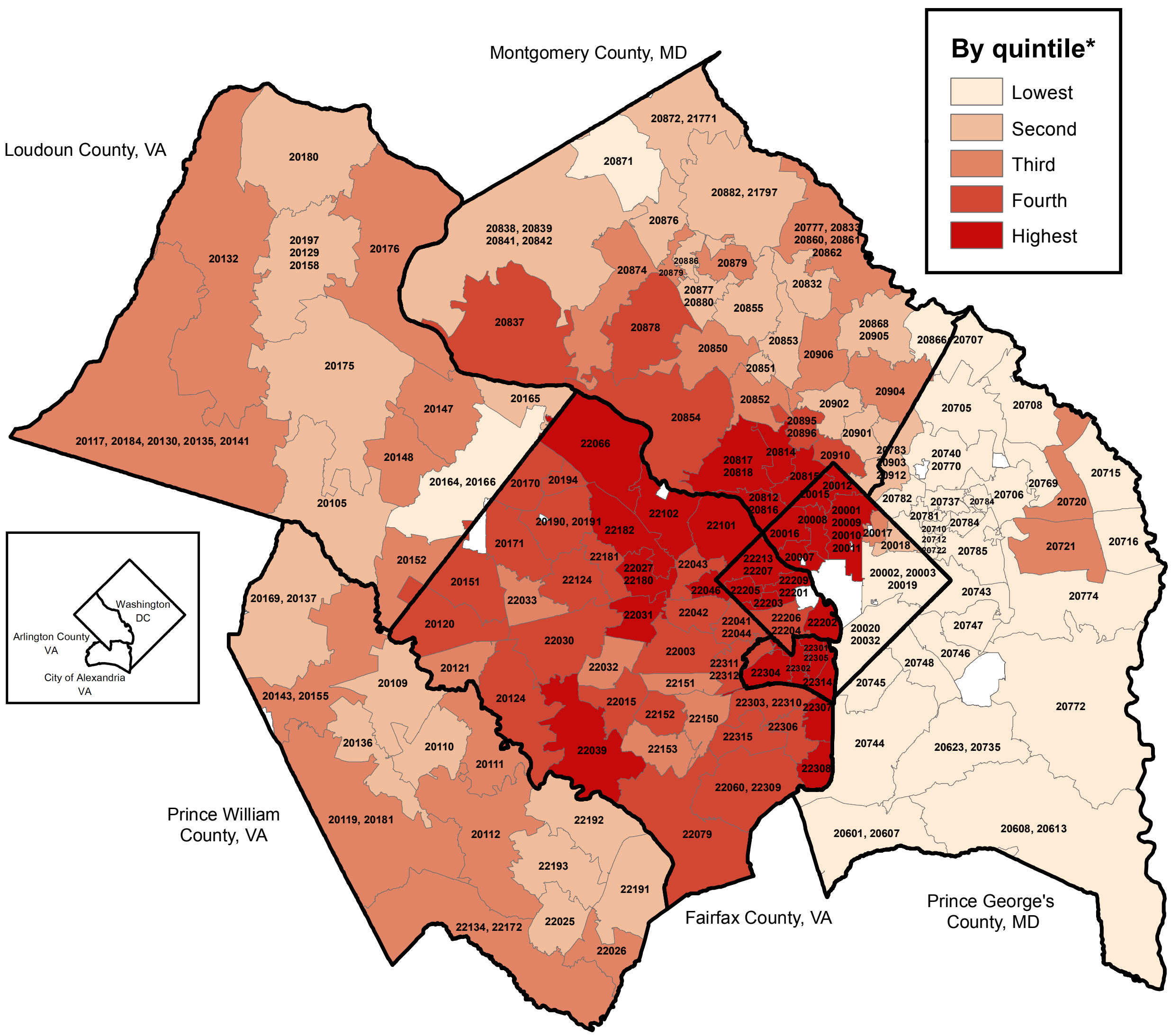
Figure 3: Average AVM, 2013:Q4



* Range of average AVM values by quintile (rounded to the nearest \$1000) is \$122,000 to \$314,000, \$314,000 to \$436,000, \$436,000 to \$562,000, \$562,000 to \$679,000, and \$679,000 to \$1,710,000.

Source: Authors' calculations using data from FNC, Inc.

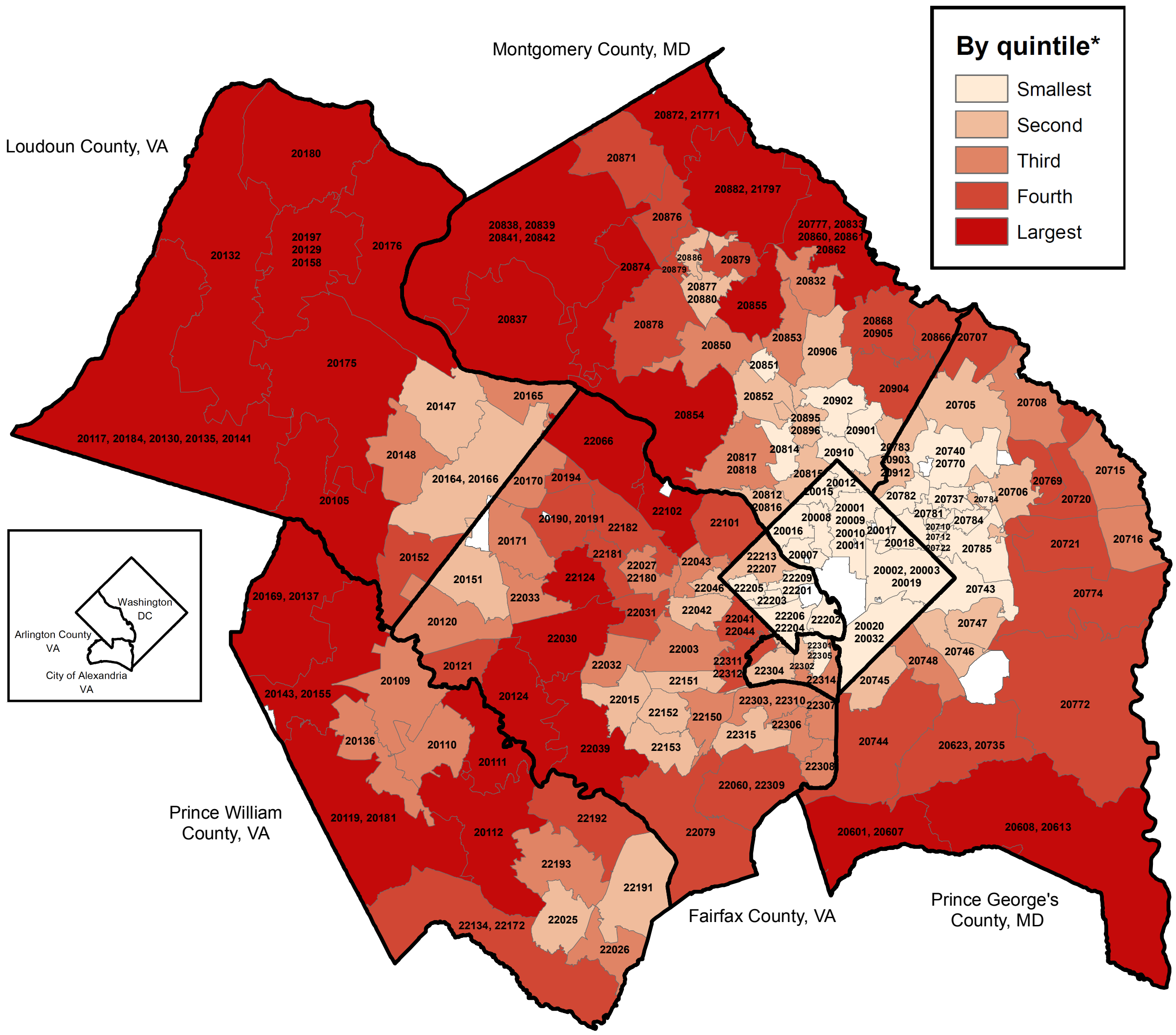
Figure 4: Average Lot Value, 2013:Q4



* Range of average lot value by quintile (rounded to the nearest \$1000) is \$24,000 to \$143,000, \$143,000 to \$211,000, \$211,000 to \$308,000, \$308,000 to \$450,000, and \$450,000 to \$1,207,000.

Source: Authors' calculations using data from FNC, Inc. and Marshall & Swift, a CoreLogic company.

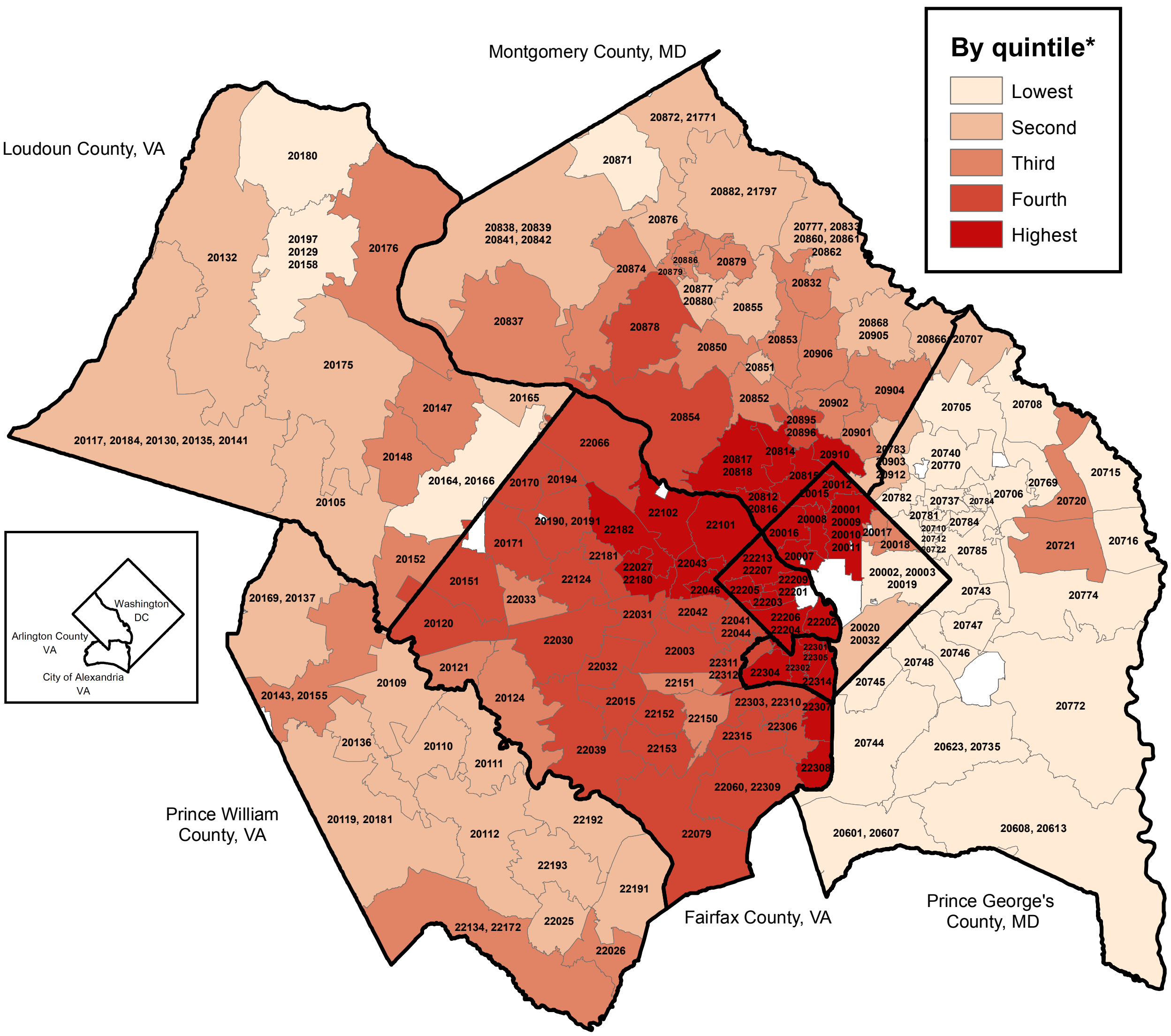
Figure 5: Average Lot Size, 2013:Q4



* Range of average lot size by quintile (rounded to the nearest 1000 square feet) is 5,000 to 9,000, 9,000 to 13,000, 13,000 to 17,000, 17,000 to 29,000, and 29,000 to 182,000.

Source: Authors' calculations using data from FNC, Inc.

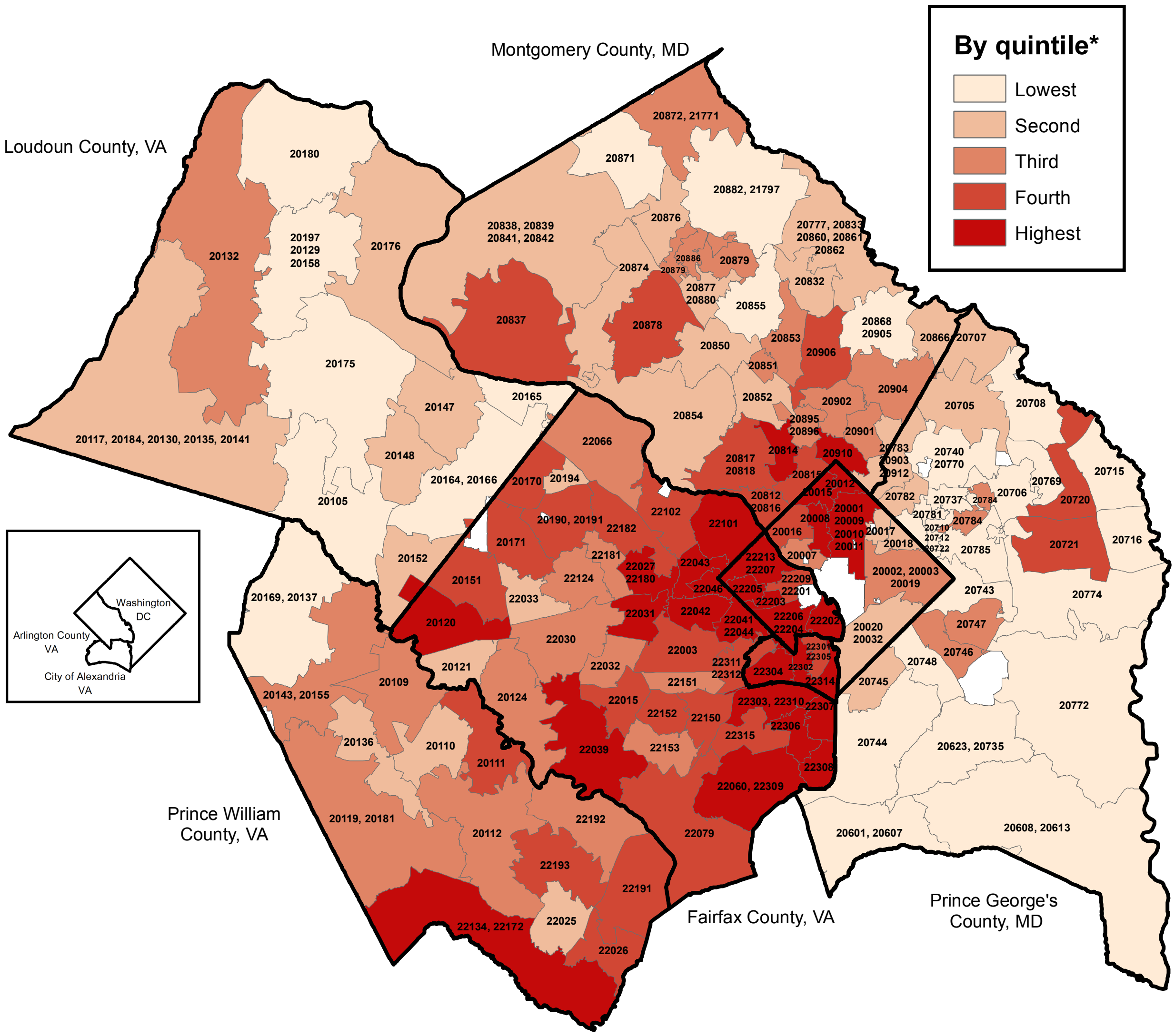
Figure 6: Average Standardized Land Price per Square Foot, 2013:Q4



* Range of average land price per square foot for a quarter-acre lot by quintile (rounded to the nearest dollar) is \$2 to \$12, \$12 to \$18, \$18 to \$25, \$25 to \$39, and \$39 to \$117.

Source: Authors' calculations using data from FNC, Inc. and Marshall & Swift, a CoreLogic company.

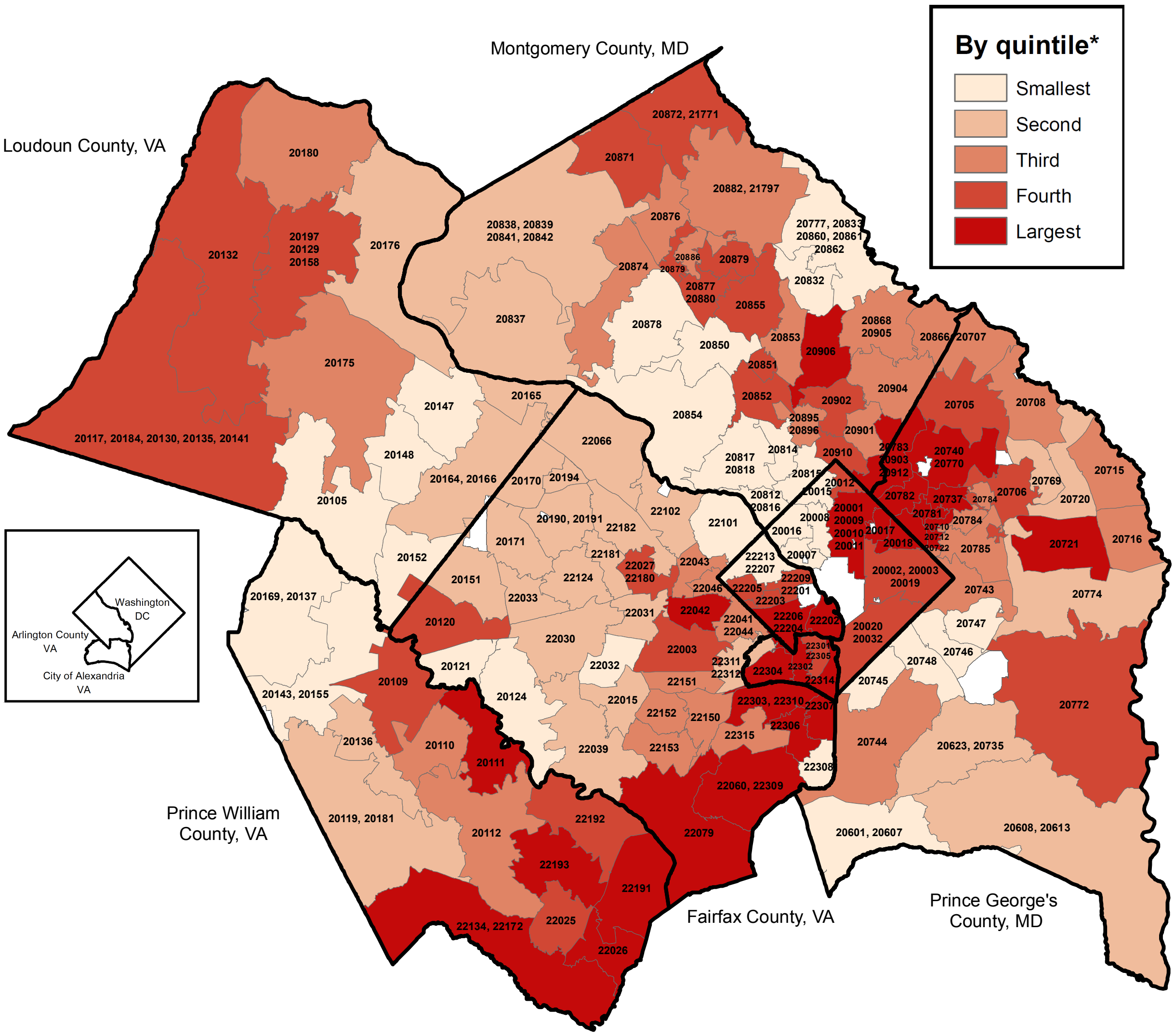
Figure 7: Average Land Share of Property Value, 2013:Q4



* Range of average land share of property value by quintile (rounded to the nearest 1%) is 7% to 37%, 37% to 48%, 48% to 58%, 58% to 72%, and 72% to 81%.

Source: Authors' calculations using data from FNC, Inc. and Marshall & Swift, a CoreLogic company.

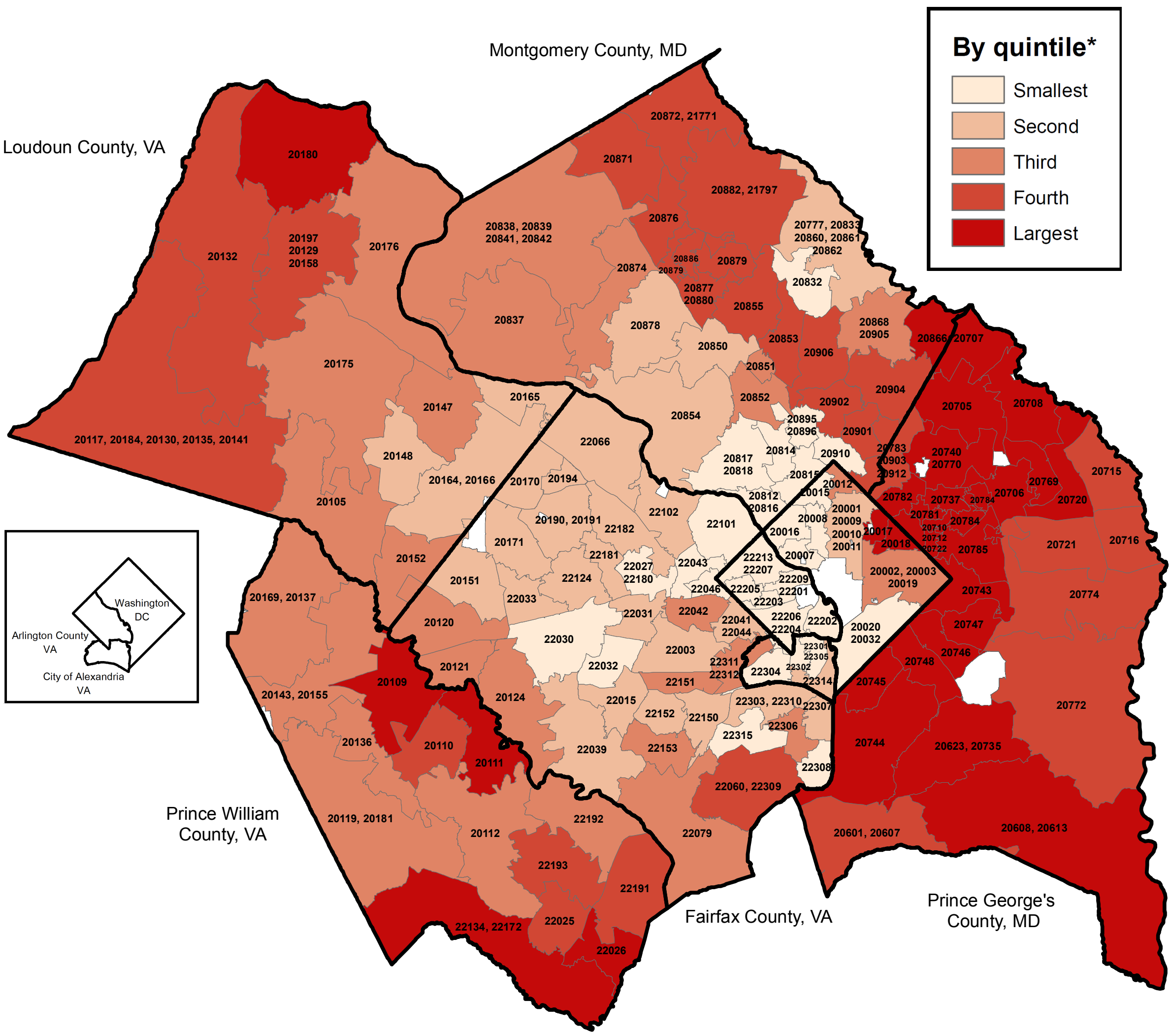
Figure 8: House Price Increase, 2000-2006



* Range of average house price increase by quintile (rounded to the nearest 1%) is 89% to 127%, 127% to 132%, 132% to 147%, 147% to 165%, and 165% to 204%.

Source: Authors' calculations using data from FNC, Inc.

Figure 9: House Price Decline, 2006-2012

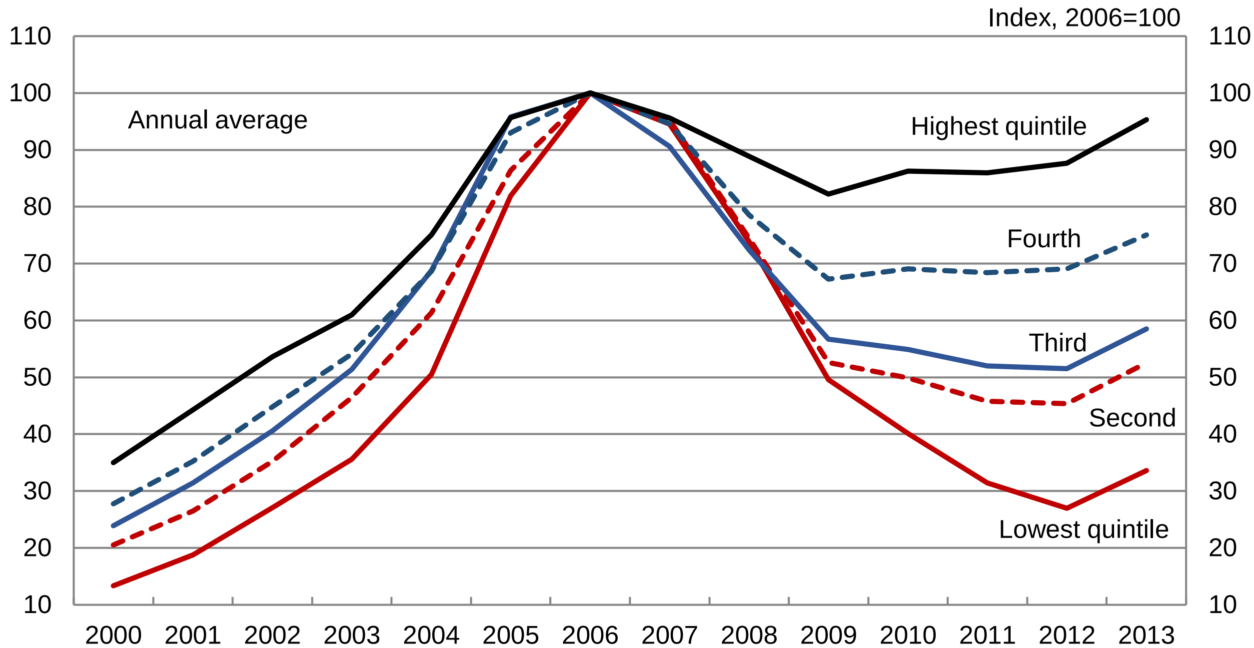


* Range of average house price decline by quintile (rounded to the nearest 1%) is less than 12%, 12% to 21%, 21% to 31%, 31% to 40%, and 40% to 51%.

Note: the 'Smallest' quintile includes zips 20007, 20008, 20016, 20815, 22201, 22203, 22205, 22207, 22209, and 22213, where house prices increased between 2006 and 2012.

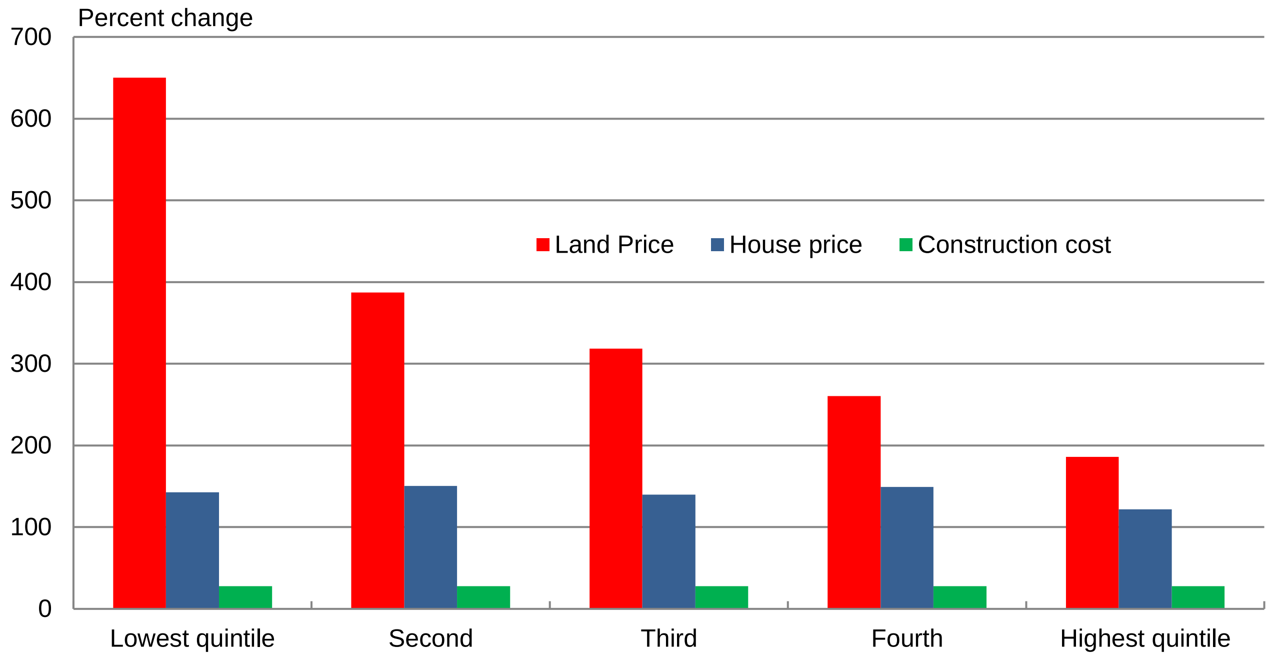
Source: Authors' calculations using data from FNC, Inc.

Figure 10: Average Land Prices: 2000-2013
(for zips grouped into quintiles by quarter-acre land price in 2000:Q1)



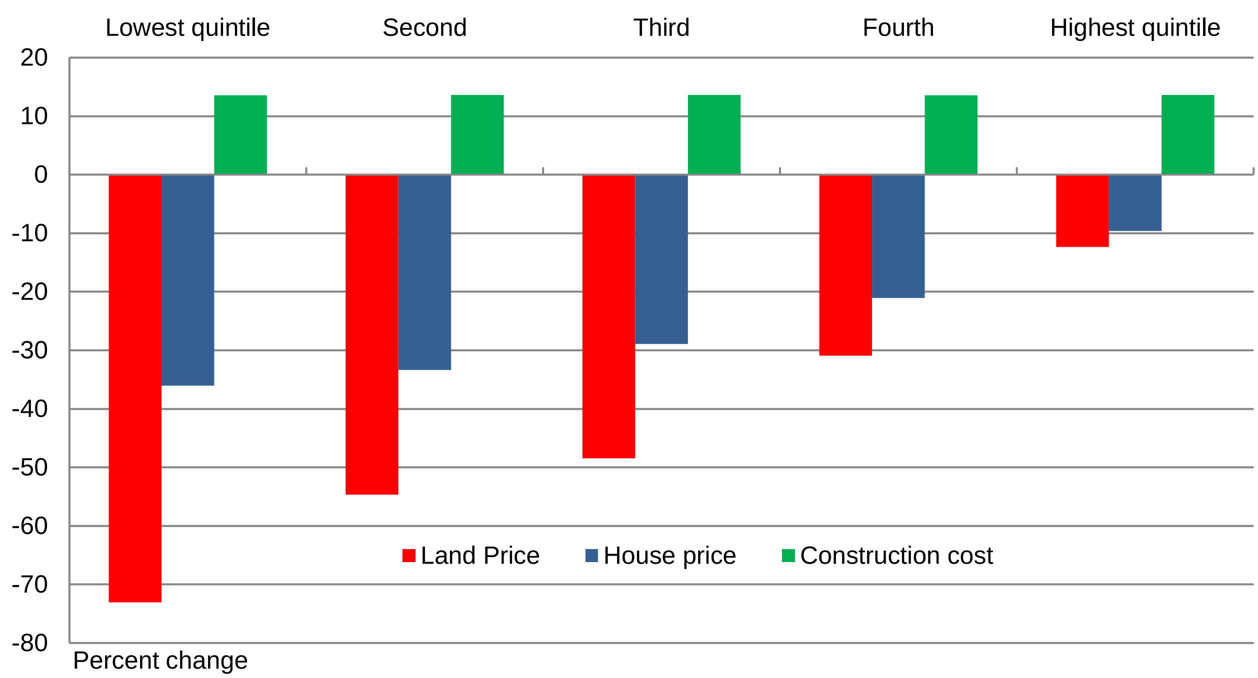
Source: Authors' calculations using data from FNC, Inc. and Marshall & Swift, a CoreLogic company.

Figure 11: Prices and Construction Costs: 2000-2006
(for zips grouped into quintiles by quarter-acre land price in 2000:Q1)



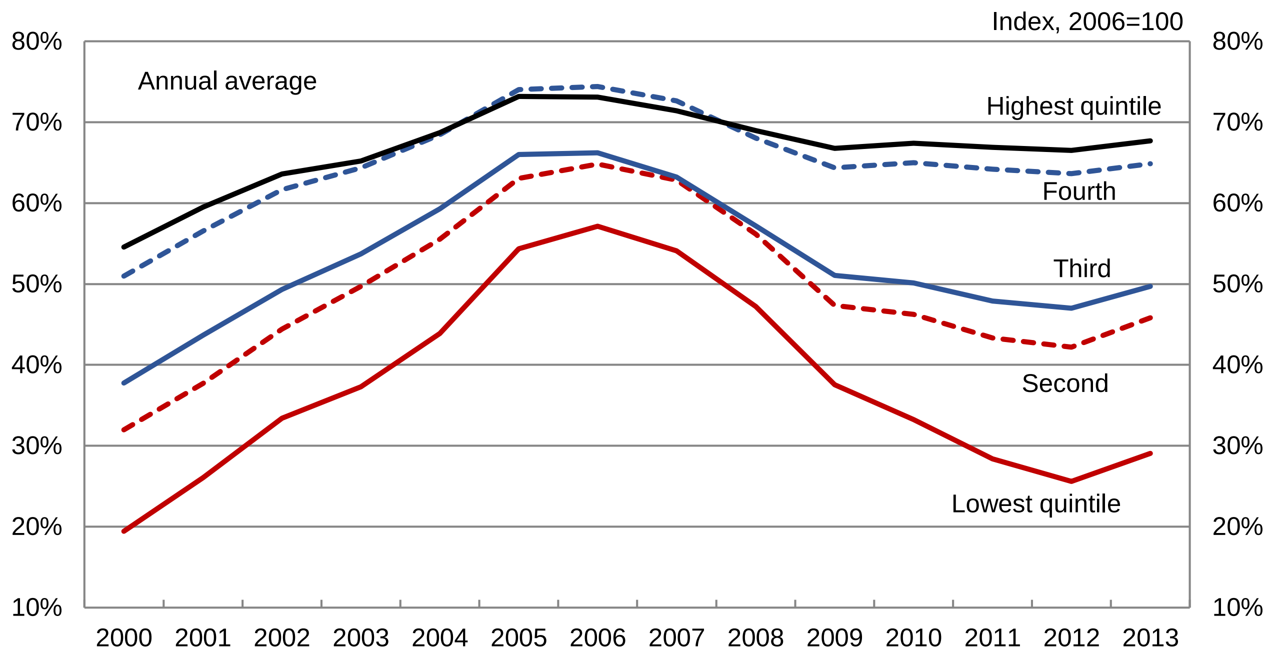
Source: Authors' calculations using data from FNC, Inc. and Marshall & Swift, a CoreLogic company.

Figure 12: Prices and Construction Costs: 2006-2012
(for zips grouped into quintiles by quarter-acre land price in 2000:Q1)



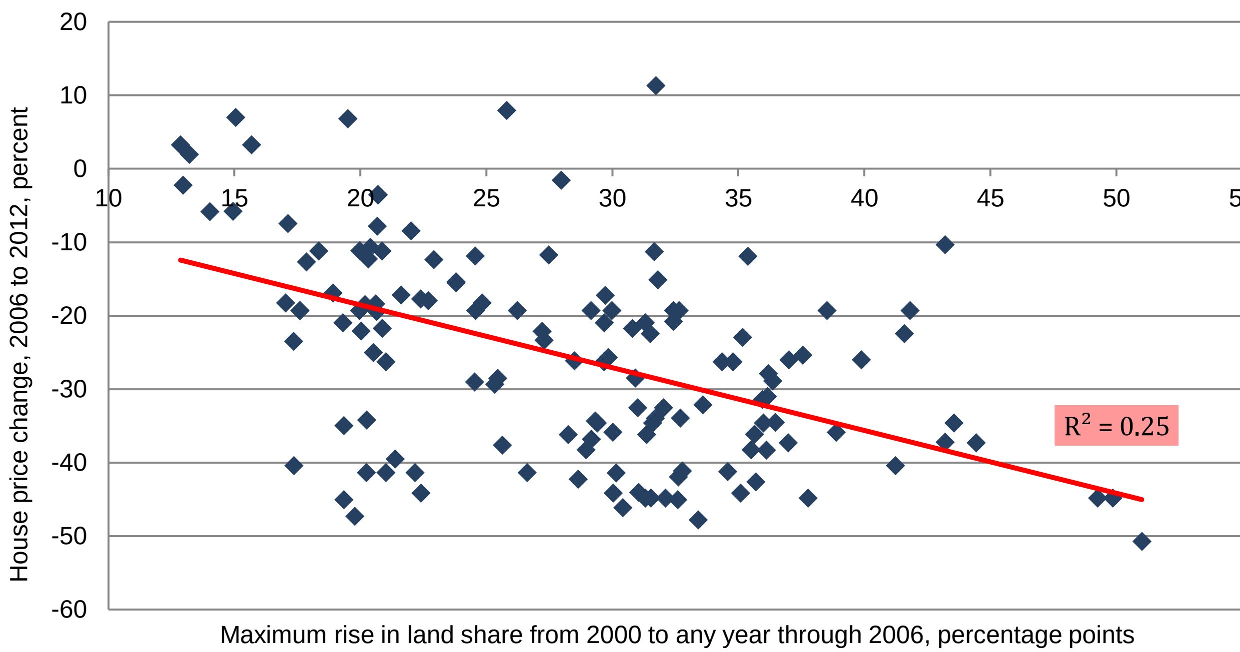
Source: Authors' calculations using data from FNC, Inc. and Marshall & Swift, a CoreLogic company.

Figure 13: Average Land Share of Home Value: 2000-2013
(for zips grouped into quintiles by quarter-acre land price in 2000:Q1)



Source: Authors' calculations using data from FNC, Inc. and Marshall & Swift, a CoreLogic company.

Figure 14: Predictive Power of Changes in Land Share



Sources: Authors' calculations using data from FNC, Inc. and Marshall & Swift, a CoreLogic company.

A Appendix

The appendix contains exhibits with supplementary information mentioned in the main part of the paper. Table A.1 provides detailed information on every zip code and zip group in the dataset. Figures A.1 through A.5 are versions of figures 10 through 14, respectively, that exclude the zip groups with adjusted land shares.

Table A.1: Detailed Information on Zip Groups

Group	Zips	County/City	# Homes	# Sales-based land values	Reference period	Type of sales used	Land Share 2000:Q1 (%)	Adjusted?
1	20007	DC	1,391	32	2003:Q2	H,L	22.2	No
2	20008	DC	2,075	22	2002:Q2	H,L	60.0	Yes
3	20012	DC	2,770	3	2004:Q3	H,L	60.0	Yes
4	20015	DC	3,870	17	2003:Q4	H,L	60.0	Yes
5	20016	DC	5,021	83	2000:Q3	H,L	51.2	No
6	20017	DC	1,389	3	NA	H,L	38.8	No
7	20018	DC	2,432	17	2001:Q3	H,L	40.4	No
8	20020, 20032	DC	2,762	81	2000:Q4	H,L	5.0	Yes
9	20001, 20009, 20010, 20011	DC	2,862	18	2003:Q2	H,L	60.0	Yes
10	20002, 20003, 20019	DC	3,076	131	2002:Q1	H,L	21.0	No
11	20814	MG	4,797	259	2000:Q3	H,L	60.0	Yes
12	20815	MG	5,910	147	2000:Q2	H,L	58.2	No
13	20832	MG	4,805	186	2000:Q2	H	23.8	No
14	20837	MG	1,439	64	2000:Q2	H	52.7	No
15	20850	MG	6,554	649	2000:Q3	H	23.9	No
16	20851	MG	3,236	1	NA	H,L	35.6	No
17	20852	MG	4,319	27	2003:Q1	H,L	15.0	No
18	20853	MG	8,247	119	2000:Q2	H	29.2	No
19	20854	MG	13,569	330	2000:Q2	H	31.9	No
20	20855	MG	3,470	26	2001:Q2	H	9.9	No
21	20871	MG	2,695	1,356	2000:Q2	H	5.0	Yes
22	20874	MG	4,831	573	2000:Q2	H,L	32.9	No
23	20876	MG	3,209	229	2000:Q2	H,L	31.1	No
24	20878	MG	9,273	637	2000:Q2	H,L	45.5	No

Jurisdictions: AR = Arlington Co., VA; AX = City of Alexandria, VA; DC = Washington, DC; FF = Fairfax Co., VA; LD = Loudoun Co., VA; MG = Montgomery Co., MD; PG = Prince George's Co., MD; PW = Prince William Co., VA.

Type of sales: H,L = home and land sales, H = home sales only.

Table A.1 – *Continued from previous page*

Group	Zips	County/City	# Homes	# Sales-based land values	Reference period	Type of sales used	Land Share 2000:Q1 (%)	Adjusted?
25	20879	MG	2,814	173	2000:Q3	H,L	34.7	No
26	20886	MG	1,643	0	NA	H,L	37.9	No
27	20901	MG	7,645	11	2003:Q2	H,L	33.6	No
28	20902	MG	9,056	31	2001:Q1	H,L	36.1	No
29	20904	MG	7,993	402	2001:Q3	H,L	45.7	No
30	20906	MG	8,678	167	2001:Q3	H,L	58.3	No
31	20910	MG	4,987	64	2000:Q2	H,L	60.0	Yes
32	20777, 20833, 20860, 20861, 20862	MG	3,337	344	2000:Q2	H	27.7	No
33	20812, 20816	MG	4,407	132	2001:Q1	H,L	45.0	No
34	20817, 20818	MG	10,851	533	2000:Q3	H,L	36.7	No
35	20838, 20839, 20841, 20842	MG	3,282	1,293	2000:Q3	H	21.1	No
36	20868, 20905	MG	5,609	303	2000:Q2	H	22.9	No
37	20872, 21771	MG	2,983	157	2000:Q3	H,L	35.8	No
38	20877, 20880	MG	2,891	70	2000:Q4	H,L	25.6	No
39	20882, 21797	MG	4,251	344	2000:Q2	H	21.7	No
40	20895, 20896	MG	6,053	148	2000:Q3	H,L	26.5	No
41	20705	PG	5,179	608	2000:Q2	H,L	36.2	No
42	20706	PG	7,426	200	2000:Q2	H,L	31.1	No
43	20708	PG	2,749	241	2000:Q2	H,L	28.3	No
44	20715	PG	8,631	394	2000:Q2	H	5.0	Yes
45	20716	PG	3,934	202	2000:Q2	H	20.9	No
46	20720	PG	5,898	1,091	2000:Q2	H	60.0	Yes
47	20721	PG	6,971	958	2000:Q3	H	60.0	Yes
48	20737	PG	2,857	78	2006:Q3	H,L	5.0	Yes
49	20743	PG	6,214	173	2000:Q2	H,L	31.2	No

Jurisdictions: AR = Arlington Co., VA; AX = City of Alexandria, VA; DC = Washington, DC; FF = Fairfax Co., VA; LD = Loudoun Co., VA; MG = Montgomery Co., MD; PG = Prince George's Co., MD; PW = Prince William Co., VA.

Type of sales: H,L = home and land sales, H = home sales only.

Table A.1 – *Continued from previous page*

Group	Zips	County/City	# Homes	# Sales-based land values	Reference period	Type of sales used	Land Share 2000:Q1 (%)	Adjusted?
50	20744	PG	13,746	897	2000:Q2	H	20.6	No
51	20745	PG	3,657	3	2002:Q2	H	48.1	No
52	20746	PG	2,953	28	2002:Q3	H	50.8	No
53	20747	PG	5,422	251	2000:Q3	H,L	52.7	No
54	20748	PG	6,232	158	2000:Q2	H,L	37.7	No
55	20769	PG	2,227	326	2001:Q2	H	5.0	Yes
56	20772	PG	10,867	1,818	2000:Q2	H	23.4	No
57	20774	PG	9,822	2,115	2000:Q2	H	18.3	No
58	20781	PG	2,050	4	2003:Q4	H,L	6.2	No
59	20782	PG	4,030	5	2000:Q4	H,L	39.9	No
60	20784	PG	5,235	2	2003:Q4	H	60.0	Yes
61	20785	PG	3,854	122	2000:Q1	H,L	32.6	No
62	20601, 20607	PG	4,223	865	2000:Q2	H	25.6	No
63	20608, 20613	PG	3,843	926	2000:Q3	H	24.1	No
64	20623, 20735	PG	11,324	1,108	2000:Q2	H	18.5	No
65	20710, 20712, 20722	PG	2,833	22	2001:Q3	H,L	28.1	No
66	20740, 20770	PG	5,694	21	2005:Q4	H,L	5.0	Yes
67	20707, 20866	PG/MG	6,623	1,389	2000:Q2	H,L	38.8	No
68	20783, 20903, 20912	PG/MG	11,018	37	2001:Q1	H,L	28.9	No
69	22202	AR	1,616	23	2002:Q1	H,L	59.9	No
70	22203	AR	1,875	17	2003:Q3	H,L	60.0	Yes
71	22205	AR	5,266	151	2001:Q1	H,L	60.0	Yes
72	22206, 22204	AR	5,108	131	2002:Q1	H,L	60.0	Yes
73	22209, 22201	AR	2,760	157	2002:Q1	H,L	45.5	No
74	22213, 22207	AR/FF	9,808	588	2000:Q2	H,L	60.0	Yes

Jurisdictions: AR = Arlington Co., VA; AX = City of Alexandria, VA; DC = Washington, DC; FF = Fairfax Co., VA; LD = Loudoun Co., VA; MG = Montgomery Co., MD; PG = Prince George's Co., MD; PW = Prince William Co., VA.

Type of sales: H,L = home and land sales, H = home sales only.

Table A.1 – *Continued from previous page*

Group	Zips	County/City	# Homes	# Sales-based land values	Reference period	Type of sales used	Land Share 2000:Q1 (%)	Adjusted?
75	22302	AX/FF	1,981	25	2001:Q2	H,L	60.0	Yes
76	22304	AX	1,854	48	2002:Q2	H,L	60.0	Yes
77	22301, 22305	AX	3,198	30	2001:Q3	H,L	41.5	No
78	22311, 22312	AX/FF	3,632	139	2000:Q3	H,L	60.0	Yes
79	20121	FF	1,572	116	2002:Q2	H,L	27.9	No
80	20124	FF	3,691	193	2000:Q2	H,L	31.7	No
81	20151	FF	4,493	527	2000:Q2	H,L	60.0	Yes
82	20170	FF	6,942	331	2000:Q3	H,L	60.0	Yes
83	20171	FF	8,241	619	2000:Q3	H,L	54.6	No
84	20194	FF	1,842	67	2000:Q3	H,L	26.3	No
85	22003	FF	10,147	172	2000:Q4	H,L	53.3	No
86	22015	FF	7,206	122	2000:Q3	H,L	41.7	No
87	22030	FF	4,587	873	2000:Q3	H,L	18.5	No
88	22031	FF	3,017	245	2000:Q2	H,L	60.0	Yes
89	22032	FF	7,069	203	2000:Q4	H,L	34.0	No
90	22033	FF	4,846	553	2000:Q3	H,L	25.8	No
91	22039	FF	5,506	154	2000:Q2	H,L	60.0	Yes
92	22042	FF	6,386	73	2001:Q3	H,L	60.0	Yes
93	22043	FF	4,160	246	2000:Q3	H,L	54.5	No
94	22046	FF	1,386	142	2000:Q2	H,L	60.0	Yes
95	22079	FF	4,443	1,836	2000:Q3	H,L	47.3	No
96	22101	FF	8,694	439	2000:Q3	H,L	59.2	No
97	22102	FF	3,284	337	2000:Q3	H,L	60.0	Yes
98	22124	FF	4,094	440	2000:Q3	H,L	33.8	No
99	22150	FF	4,791	278	2000:Q2	H,L	49.6	No

Jurisdictions: AR = Arlington Co., VA; AX = City of Alexandria, VA; DC = Washington, DC; FF = Fairfax Co., VA; LD = Loudoun Co., VA; MG = Montgomery Co., MD; PG = Prince George's Co., MD; PW = Prince William Co., VA.

Type of sales: H,L = home and land sales, H = home sales only.

Table A.1 – *Continued from previous page*

Group	Zips	County/City	# Homes	# Sales-based land values	Reference period	Type of sales used	Land Share 2000:Q1 (%)	Adjusted?
100	22151	FF	4,546	40	2003:Q4	H,L	38.2	No
101	22152	FF	4,669	126	2001:Q2	H,L	50.8	No
102	22153	FF	6,185	253	2000:Q2	H,L	41.0	No
103	22181	FF	3,157	131	2000:Q3	H,L	40.4	No
104	22182	FF	6,778	474	2000:Q3	H,L	44.2	No
105	22306	FF	3,362	134	2001:Q1	H,L	59.4	No
106	22308	FF	4,570	72	2001:Q1	H,L	60.0	Yes
107	22315	FF	2,457	320	2000:Q3	H,L	37.5	No
108	20190, 20191	FF	3,841	8	2001:Q2	H,L	60.0	Yes
109	22027, 22180	FF	6,519	459	2000:Q4	H,L	54.7	No
110	22041, 22044	FF	3,482	121	2001:Q4	H,L	60.0	Yes
111	22060, 22309	FF	4,928	253	2000:Q3	H,L	60.0	Yes
112	22303, 22310	FF	6,468	311	2000:Q2	H,L	60.0	Yes
113	22314, 22307	FF/AX	2,738	40	2001:Q4	H,L	60.0	Yes
114	20120	FF/LD	6,373	816	2000:Q3	H,L	60.0	Yes
115	22066	FF/LD	5,575	378	2000:Q2	H,L	31.9	No
116	20105	LD	2,798	1,506	2001:Q3	H	7.8	No
117	20147	LD	7,672	2,248	2000:Q3	H,L	24.3	No
118	20148	LD	6,553	4,375	2000:Q3	H,L	20.6	No
119	20152	LD	5,553	2,741	2000:Q3	H	14.4	No
120	20165	LD	5,073	300	2000:Q3	H	12.5	No
121	20175	LD	6,012	1,393	2000:Q3	H	11.4	No
122	20176	LD	8,203	2,798	2000:Q3	H	22.6	No
123	20180	LD	2,234	786	2000:Q3	H,L	37.0	No
124	20132	LD	4,274	1,034	2000:Q3	H	5.0	Yes

Jurisdictions: AR = Arlington Co., VA; AX = City of Alexandria, VA; DC = Washington, DC; FF = Fairfax Co., VA; LD = Loudoun Co., VA; MG = Montgomery Co., MD; PG = Prince George's Co., MD; PW = Prince William Co., VA.

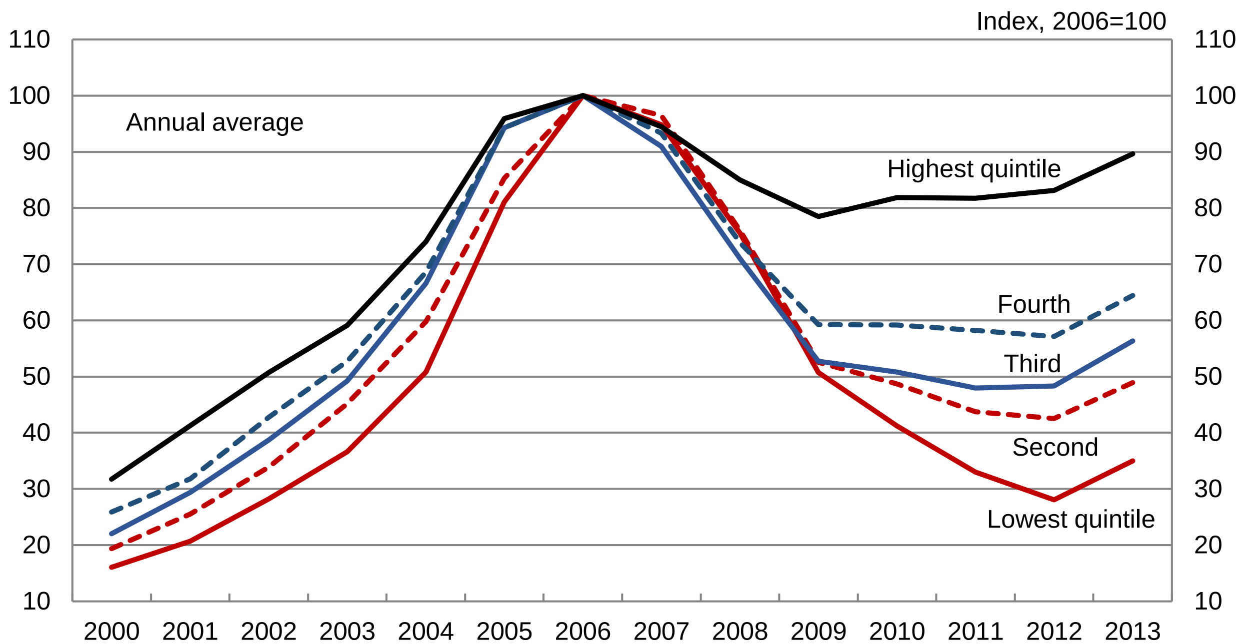
Type of sales: H,L = home and land sales, H = home sales only.

Table A.1 – *Continued from previous page*

Group	Zips	County/City	# Homes	# Sales-based land values	Reference period	Type of sales used	Land Share 2000:Q1 (%)	Adjusted?
125	20117, 20184, 20130, 20135, 20141	LD	2,550	552	2001:Q3	H	5.0	Yes
126	20164, 20166	LD	6,826	278	2002:Q4	H,L	5.0	Yes
127	20197, 20129, 20158	LD	1,934	419	2000:Q3	H,L	7.3	No
128	20109	PW	2,292	209	2002:Q3	H	60.0	Yes
129	20110	PW	1,769	375	2000:Q3	H,L	32.9	No
130	20111	PW	4,117	788	2000:Q3	H	50.1	No
131	20112	PW	8,802	2,262	2000:Q2	H,L	44.0	No
132	20136	PW	7,264	4,233	2000:Q3	H,L	32.5	No
133	22026	PW	2,395	1,144	2000:Q3	H,L	46.4	No
134	22191	PW	7,876	1,832	2000:Q3	H,L	45.4	No
135	22192	PW	7,506	1,138	2000:Q2	H	27.4	No
136	22193	PW	15,076	2,436	2000:Q3	H	46.3	No
137	22025	PW	3,757	1,053	2000:Q2	H	38.2	No
138	20119, 20181	PW	2,198	339	2001:Q4	H,L	44.4	No
139	20143, 20155	PW	8,411	5,051	2000:Q3	H,L	39.2	No
140	20169, 20137	PW	6,459	3,094	2000:Q3	H	15.3	No
141	22134, 22172	PW	1,855	811	2002:Q3	H,L	60.0	Yes

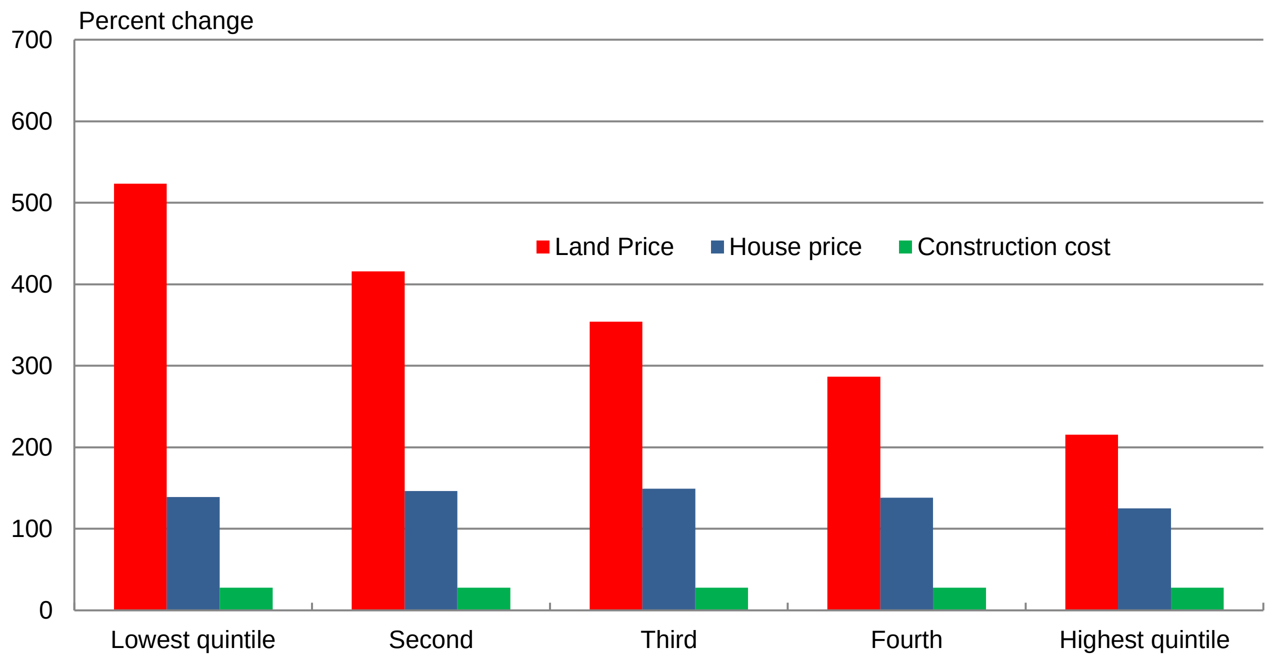
Sources: Authors' calculations using data from FNC, Inc. and Marshall & Swift, a CoreLogic company.

Figure A.1: Average Land Prices: 2000-2013
Excluding Zip Groups with Adjusted Land Shares
(remaining zips grouped into quintiles by quarter-acre land price in 2000:Q1)



Source: Authors' calculations using data from FNC, Inc. and Marshall & Swift, a CoreLogic company.

Figure A.2: Prices and Construction Costs: 2000-2006
Excluding Zip Groups with Adjusted Land Shares
(remaining zips grouped into quintiles by quarter-acre land price in 2000:Q1)

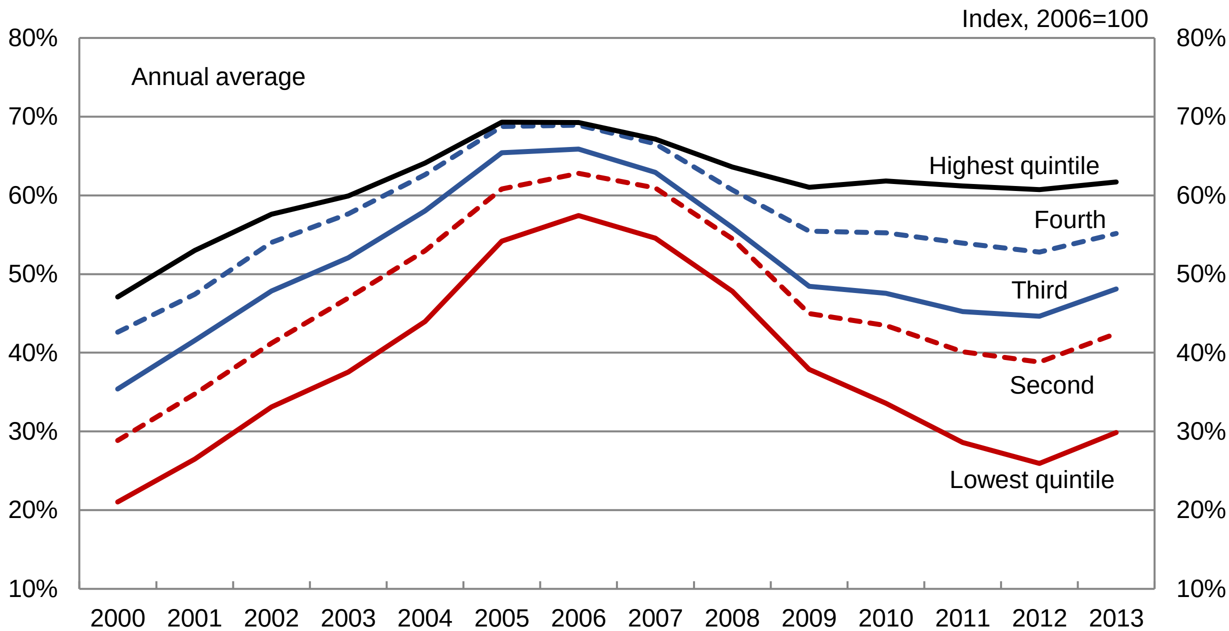


Source: Authors' calculations using data from FNC, Inc. and Marshall & Swift, a CoreLogic company.

Figure A.3: Prices and Construction Costs: 2006-2012
Excluding Zip Groups with Adjusted Land Shares
(remaining zips grouped into quintiles by quarter-acre land price in 2000:Q1)

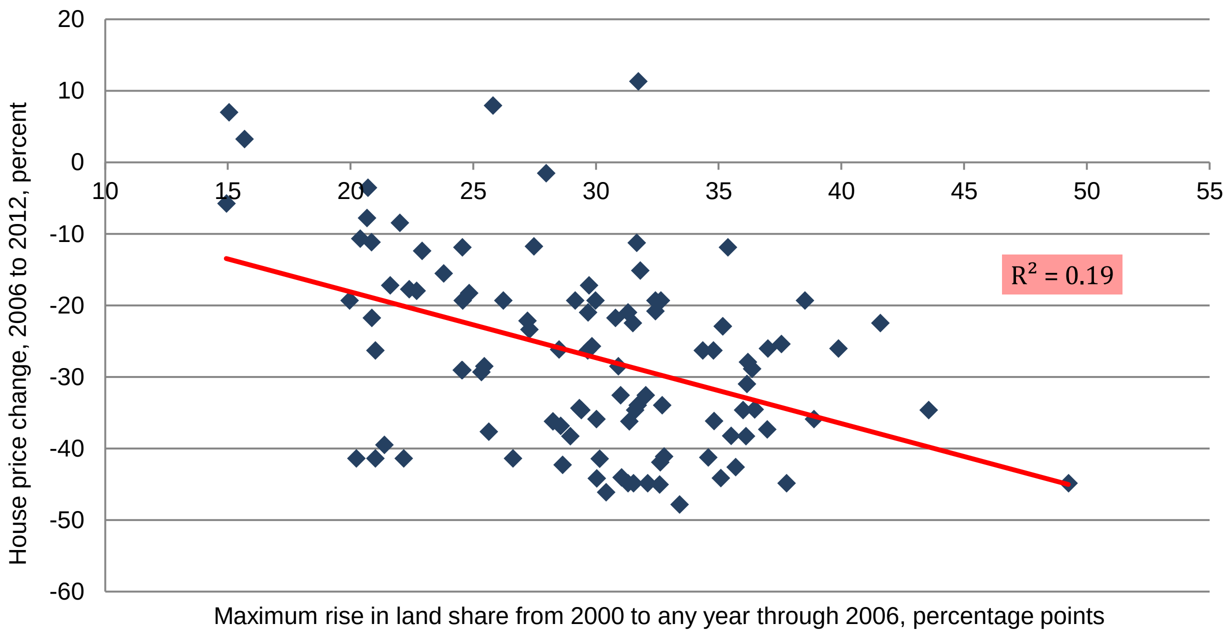
Source: Authors' calculations using data from FNC, Inc. and Marshall & Swift, a CoreLogic company.

Figure A.4: Average Land Share of Home Value: 2000-2013
Excluding Zip Groups with Adjusted Land Shares
(remaining zips grouped into quintiles by quarter-acre land price in 2000:Q1)



Source: Authors' calculations using data from FNC, Inc. and Marshall & Swift, a CoreLogic company.

Figure A.5: Predictive Power of Changes in Land Share
Excluding Zip Groups with Adjusted Land Shares



Source: Authors' calculations using data from FNC, Inc. and Marshall & Swift, a CoreLogic company.